

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:

January 8, 2001, 09:20:44 ; Search time 2611.99 Seconds  
(Without alignments)  
2315.928 Million cell updates/sec

Title: us-09-553-431-3

Perfect score: 1182

Sequence: 1 aagctgatactgcacactcc.....tcaaaaaaaaaaaaaaaaaa 1182

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched:

1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters:

2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank1:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_ba3:  
4: gb\_ba4:  
5: gb\_ba5:  
6: gb\_ba6:  
7: gb\_ba7:  
8: gb\_ba8:  
9: gb\_ba9:  
10: gb\_ba10:  
11: gb\_ba11:  
12: gb\_ba12:  
13: gb\_ba13:  
14: gb\_ba14:  
15: gb\_ba15:  
16: gb\_ba16:  
17: gb\_ba17:  
18: gb\_ba18:  
19: gb\_ba19:  
20: gb\_ba20:  
21: gb\_ba21:  
22: gb\_ba22:  
23: gb\_ba23:  
24: gb\_ba24:  
25: gb\_ba25:  
26: gb\_ba26:  
27: gb\_ba27:  
28: gb\_ba28:  
29: gb\_ba29:  
30: gb\_ba30:  
31: gb\_ba31:  
32: gb\_ba32:  
33: gb\_ba33:  
34: gb\_ba34:  
35: gb\_ba35:  
36: gb\_ba36:  
37: gb\_ba37:  
38: gb\_ba38:  
39: gb\_ba39:  
40: gb\_ba40:  
41: gb\_ba41:  
42: gb\_ba42:  
43: gb\_ba43:

44: em\_htg8:  
45: em\_htg9:  
46: em\_htg10:  
47: em\_htg11:  
48: em\_htg12:  
49: em\_htg13:  
50: em\_htg14:  
51: em\_htg15:  
52: em\_htg16:  
53: em\_htg17:  
54: em\_htg18:  
55: em\_htg19:  
56: em\_htg20:  
57: em\_htg21:  
58: em\_htg22:  
59: em\_htg23:  
60: em\_htg24:  
61: em\_htg25:  
62: em\_htg26:  
63: em\_htg27:  
64: em\_htg28:  
65: em\_htg29:  
66: em\_htg30:  
67: em\_htg31:  
68: em\_htg32:  
69: em\_htg33:  
70: em\_htg34:  
71: em\_htg35:  
72: em\_htg36:  
73: em\_htg37:  
74: em\_htg38:  
75: em\_htg39:  
76: em\_htg40:  
77: em\_htg41:  
78: em\_htg42:  
79: em\_htg43:  
80: em\_htg44:  
81: em\_htg45:  
82: em\_htg46:  
83: em\_htg47:  
84: em\_htg48:  
85: em\_htg49:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1164	98.5	1164	7	AF251019
2	455.8	38.6	65958	6	AB009056
3	454	38.4	981	6	AB030278
4	205.8	17.4	118360	6	AF16114
5	185.4	15.7	150613	6	AB001684
6	163.8	13.9	200799	6	AF137379
7	159.6	13.5	22085	33	PM2245645
8	157.8	13.4	121524	6	AF041468
9	157.8	13.4	121524	6	AF041468
10	157.8	13.4	121524	6	AF041468
11	157.8	13.4	121524	6	AF041468
12	157.8	13.4	121524	6	AF041468
13	157.8	13.4	121524	6	AF041468
14	157.8	13.4	121524	6	AF041468
15	157.8	13.4	121524	6	AF041468
16	157.8	13.4	121524	6	AF041468
17	157.8	13.4	121524	6	AF041468
18	157.8	13.4	121524	6	AF041468
19	157.8	13.4	121524	6	AF041468
20	157.8	13.4	121524	6	AF041468
21	157.8	13.4	121524	6	AF041468

## ALIGNMENTS

FEATURES

CDS

BASE COUNT  
ORIGIN

Query Match	98.5%;	Score 1164;	DB 7;	Length 1164;
Best Local Similarity	100.0%;	Pred. NO. 7.3e-256;		
Matches 1164; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

98.58; Score 1164; DB 7; Length 1164;

```
;; Length 1164;
```

QY	1	aagcttgatatacgaaactccataactatgac	ttctcttcttcttcgcggagatgaaatccct	60
Db	1	AAGGTTGATATCGCAACCTCCATACGATCT	TTCTTCTTCTTCGCGGAGATGACATCCCT	60
QY	61	gaggtttcttaacagaaccctcaacttgatg	ctatcgtcatccattctcccaaatlcaatccct	120
Db	61	GAGGTTTCTTAACAGAACCCCTCACTGTGAT	GTGCTCATCTCCATCTTCCCACTTAAATCCCT	120
QY	121	acacaaaacccctaaactaaaccacaacca	aacctctcccaaaagccacacccaatctgc	180
Db	121	ACACAAAACCCCTAACTAAACCAACCAAAA	CCCTACCCAAAGCCACACCAATTCGGTTC	180
QY	181	gctctcttaataacatctgcacaacagagc	ctgcgcggagacatccacgagctgcgcgaat	240
Db	181	CGCTCTTCAATACATGCAATGCAAAACCA	GCAGCTCGCGGAGACACTCCAGATGCTGC	240
QY	241	cgaagccgaagcttggcttcaagtaacctg	acatctctctctcgcgtctcgaaaacgcgt	300
Db	241	CGAGCGCGGACGGTGGCTTACGTAACCTG	ATCTCTTCTGCGTCTGAAAAACCGCCTCA	300
QY	301	ttacacgcgtcgttgaagttctcaacgagc	atctgaagctcgacccaaccccttaagtga	360
Db	301	TTACACCGTCTGCTGAGATTCTCACGGGA	TTGCTACGATCCACCAACCCCTAGTTCGT	360
QY	361	taaacgctgtcaaatctcgaaatctgtgat	ctatcttcaaaaacctgataaatctgctt	420
Db	361	TAAACGCTGTCAAAATTTGGAATGCTTGT	GTATTCAAAACCTCTGAGTCCAAAATTTG	420
QY	421	aggatcttggggaaagccttagtttgctgt	atgatgatctaaagaatgaagcaagaagttg	480
Db	421	AGGATTTGGGGGAAAGCCTTAGTTTGGCT	GTATGATTTAAAGATTAAGCAAGAAGTTG	480
QY	481	cccgattcttaactataagatctgctctgc	agataltgagtcgcggatctcaacgcgc	540
Db	481	CCCGATTCTTAACCTATAGATTGCTCGAG	GTATGATTTGATTCGCCGTTCAATACGCC	540
QY	541	tacacgcgctaaagaaacgctatttagttt	aaacactgtatatactgcatltagagaatgc	600
Db	541	TACACCGGCTTACCAAGCCGATTAAGTTA	GCATACACTGATATTACTGCTTAGAGATGC	600
QY	601	agatagagttacaagctctgcttgaatgtg	atgaatlaagagatataaaatgatgtgtaa	660
Db	601	AGATAGAGTTACAGGCTTGCTTGAATGTGA	TGATTAAGGAATTTAAATGATTTGTCAA	660
QY	661	cgagatctgaactgtatgttaagaagggtg	aaagatatatagtlcaggtctctgaatga	720
Db	661	CAGAGTTGAAGTGTATGTGATTAAGCGGT	GAAGATATGATCTGATCTTGATCTTCAAA	720
QY	721	gagtgtggagattgcatctgtgtgagtata	cccgaaggattcgaagiatattcgagatcga	780
Db	721	GAGTGTGGGATTTGCTATTTGTGAGTAAT	CCGAGAGATTCGAMGTATTTGGAGTTCGAA	780
QY	781	tagaaggttccgcttggcttgaatgaacgc	ccctccgactcttagcaggatctgacattga	840
Db	781	TAGAAGGTTTCCCTTGCTGTGTTAACAAG	CCCTCCGACTCTTAGCAGATTTGCATTGAG	840
QY	841	gagctgcctggagaattggttgaacaagat	atagaatgaaggctgtgattgttgaaagaac	900
Db	841	GAGCTGCTTGAGATTTGGTTGACCAAGAT	ATGCAATGAAGCTGTGATGGTGGAGAAAC	900
QY	901	taaaaagaaggagattctctcgttcttttg	gaagttagtgatcgaaatccgttgaatcgt	960
Db	901	TAAAAGAGGGATTTTTCGTTTTCGTTT	TGAGAGTTAGTATCGAATTCGTTGAATCG	960
QY	961	agttggatttggcttggatgagaatgctgt	ctgttctgtctcaatgaaagctgcatatgt	1020
Db	961	AGTTGGATTGCTTGGTGGAGAAAGTGTCT	TGTCTTCAATGATAGAGCGCTGATATGTG	1020
QY	1021	tcacttaaatgatatgtgtacagtaagctga	atgaagattgttttaattcagtttga	1080
Db	1021	TCACCTTAAATGTTATGTGTATGATAGTA	AGCTGATGAAGATTTGTTTAAATTCAGT	1080

protein"  
join(8240..8326,8416..8695,8789..9161,9246..9406,  
9571..9698,9776..10024,10109..10399)  
/note="gene\_id:MZF18.16"  
/codon\_start=1  
/evidence=not\_experimental  
/product="(+) -delta-cadnine synthase"  
/protein\_id="BAB08719.1"  
/db\_xref="GI:97588220"  
/translation="MGESEVRLPADPEPANIWEDPLTFSKSDLGTEFPKRSHTLKE  
LVAEPMKSNAPNTENTKIFDVALCRLGVSYPHEKDIYEDLQKSPCLDPEKRWKHTLKE  
ITLVGILTFVQFGRKSLSSDSEKEDKEDNGKFRGHLVTAQVAGSLYEAQKQTHGSG  
DIDDELAVSRHLEKIDISSSPSHLARIKNAKHAPHKISRTETQISLYEYESE  
CEPTLLEAFKIDFNNLILITLALDDPELACVTRHMKPEKSKVTYIRHRTIYALMAGTQV  
EPOYLSOAIVITFMNLIITLALDDPELIPDSMKKPIYATVEYQYDLDELKEEGS  
GCGFHKAKSIQTANGYQOEAKMLKDDYLATPEYKEMALLDSQYVALLIMTFVPMID  
VAKIDAFELSLSPKIRVASEIISRTDDISSEFHKRHHVATGIDCYMOQFSEKRE  
RAVEWGNIVISQAMKDLNDELMPHPPPLRLVRLNLSRIDVFYRYQDAYTPKLL  
KHHYSLLEITPT  
11597..12883  
/note="gene\_id:MZF18.15"  
/codon\_start=1  
/evidence=not\_experimental  
/product="acetyl-coA:benzylalcohol acetyltransferase-like  
protein"  
/protein\_id="BAB08720.1"  
/db\_xref="GI:97589221"  
/translation="MKLELSEKVIKAPSPNHLQTLSLSPDOPLRSPVSAIFPYDD  
HSNDEQRLKSLISQTLSEPLVLAGQIKDQVATKDENGALFTSRAPIELISDFLRN  
SPADLDIQKIVSDVDAIDPEWPLHLYVIEFFKDFGFAVSSGIKIDASLSEFLRN  
SKTASGAGQIVSDVDAIDPEWPLHLYVIEFFKDFGFAVSSGIKIDASLSEFLRN  
NRASGGRVQAQRIEISITALLRCIKRAGSKSGKQWEPAITQMDLAPRVSSILP  
KHAIGNFPLPLKSESESKMEIETYSKLOKQOELMLIRNDSDEAKSVEAKRI  
ASAMLSLELSEMEIYVAVSSRMSEFYKEMFGMKRQVWVAAPDSVDKQTVLMSKD  
SECEVRLVTPEDMKKREHDSSELYVATISPSSTLI"  
complement(join(13448..13580,13695..14419,14516..14753,  
14956..15283,15400..15634,15951..16167,16240..16311,  
16400..16553))  
/note="gene\_id:MZF18.14"  
/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
/protein\_id="BAB08721.1"  
/db\_xref="GI:97588222"  
/translation="MGNMSLVKLTLMVLLFLGNILVIMISTNLFKSKMPTLSKYTN  
TTPYRQCPNTVLLTLPKMFIALVSCVYLRIRKLSQSYLYEEREMKLKIMKRVIN  
MNPGLIYATBELTSLRLVALLVMSLYNLYLYVIEGHANNDQAKIMQAFRAGELI  
GYVNCILAELEFPYTRASTLTPLVGLTSESSIKYIIMIGHVNSPCLVHTVFLIY  
AMIKMLKEFAMNPYVNLNAGTIAVIGIAMWVNSLPSFRKKEFEIFYTHLGYLY  
IIVFVTHIGSDMFCMLINIFLFEFDYRLRTQSKRSLVARSAILPSNLLETFSK  
SLLDLESTPLSLFHPASTSIQMHGPELTSSNLEKOTLSVIRIROSQWOKLYTHLS  
DVLIDSEKHYHDLAFLDITPLDMSADISRLNRIEAYITREDKPEPTDQHTLLO  
TYKREPORLPDSIPVIGAPNANFLVGLIYSLSFVFLILGIVTRYIYIVPHNTGSL  
YNSYERILMLPGSLCAFISSVYVFLMRKONKGGKDEFKNOQVSEKOPTSPSGS  
KAKCSOLAN(LHLEAS)FN"  
complement(join(20835..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.13"  
/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
/protein\_id="BAB08722.1"  
/db\_xref="GI:97588223"  
/translation="MGNMSLVKLMVLLFLGNIFVIMISTNLFKSKMPTLSKYLY  
TTPYRQCPNTVLLTLPKMFIALVSCVYLRIRKLSQSYLYEEREMKLKIMKRVIN  
MNPGLIYATBELTSLRLVALLVMSLYNLYLYVIEGHANNDQAKIMQAFRAGELI  
GYVNCILAELEFPYTRASTLTPLVGLTSESSIKYIIMIGHVNSPCLVHTVFLIY  
AMIKMLKEFAMNPYVNLNAGTIAVIGIAMWVNSLPSFRKKEFEIFYTHLGYLY  
IIVFVTHIGSDMFCMLINIFLFEFDYRLRTQSKRSLVARSAILPSNLLETFSK  
SLLDLESTPLSLFHPASTSIQMHGPELTSSNLEKOTLSVIRIROSQWOKLYTHLS  
DVLIDSEKHYHDLAFLDITPLDMSADISRLNRIEAYITREDKPEPTDQHTLLO  
TYKREPORLPDSIPVIGAPNANFLVGLIYSLSFVFLILGIVTRYIYIVPHNTGSL  
YNSYERILMLPGSLCAFISSVYVFLMRKONKGGKDEFKNOQVSEKOPTSPSGS  
KAKCSOLAN(LHLEAS)FN"  
complement(join(20835..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.12"  
/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
/protein\_id="BAB08723.1"  
/db\_xref="GI:97588224"  
/translation="MGNMSLVKLMVLLFLGNIFVIMISTNLFKSKMPTLSKYLY  
TTPYRQCPNTVLLTLPKMFIALVSCVYLRIRKLSQSYLYEEREMKLKIMKRVIN  
MNPGLIYATBELTSLRLVALLVMSLYNLYLYVIEGHANNDQAKIMQAFRAGELI  
GYVNCILAELEFPYTRASTLTPLVGLTSESSIKYIIMIGHVNSPCLVHTVFLIY  
AMIKMLKEFAMNPYVNLNAGTIAVIGIAMWVNSLPSFRKKEFEIFYTHLGYLY  
IIVFVTHIGSDMFCMLINIFLFEFDYRLRTQSKRSLVARSAILPSNLLETFSK  
SLLDLESTPLSLFHPASTSIQMHGPELTSSNLEKOTLSVIRIROSQWOKLYTHLS  
DVLIDSEKHYHDLAFLDITPLDMSADISRLNRIEAYITREDKPEPTDQHTLLO  
TYKREPORLPDSIPVIGAPNANFLVGLIYSLSFVFLILGIVTRYIYIVPHNTGSL  
YNSYERILMLPGSLCAFISSVYVFLMRKONKGGKDEFKNOQVSEKOPTSPSGS  
KAKCSOLAN(LHLEAS)FN"  
complement(join(20835..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.11"  
/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
/protein\_id="BAB08724.1"  
/db\_xref="GI:97588225"  
/translation="MGNMSLVKLMVLLFLGNIFVIMISTNLFKSKMPTLSKYLY  
TTPYRQCPNTVLLTLPKMFIALVSCVYLRIRKLSQSYLYEEREMKLKIMKRVIN  
MNPGLIYATBELTSLRLVALLVMSLYNLYLYVIEGHANNDQAKIMQAFRAGELI  
GYVNCILAELEFPYTRASTLTPLVGLTSESSIKYIIMIGHVNSPCLVHTVFLIY  
AMIKMLKEFAMNPYVNLNAGTIAVIGIAMWVNSLPSFRKKEFEIFYTHLGYLY  
IIVFVTHIGSDMFCMLINIFLFEFDYRLRTQSKRSLVARSAILPSNLLETFSK  
SLLDLESTPLSLFHPASTSIQMHGPELTSSNLEKOTLSVIRIROSQWOKLYTHLS  
DVLIDSEKHYHDLAFLDITPLDMSADISRLNRIEAYITREDKPEPTDQHTLLO  
TYKREPORLPDSIPVIGAPNANFLVGLIYSLSFVFLILGIVTRYIYIVPHNTGSL  
YNSYERILMLPGSLCAFISSVYVFLMRKONKGGKDEFKNOQVSEKOPTSPSGS  
KAKCSOLAN(LHLEAS)FN"  
complement(join(20835..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.10"  
/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
/protein\_id="BAB08725.1"  
/db\_xref="GI:97588226"  
/translation="MGNMSLVKLMVLLFLGNIFVIMISTNLFKSKMPTLSKYLY  
TTPYRQCPNTVLLTLPKMFIALVSCVYLRIRKLSQSYLYEEREMKLKIMKRVIN  
MNPGLIYATBELTSLRLVALLVMSLYNLYLYVIEGHANNDQAKIMQAFRAGELI  
GYVNCILAELEFPYTRASTLTPLVGLTSESSIKYIIMIGHVNSPCLVHTVFLIY  
AMIKMLKEFAMNPYVNLNAGTIAVIGIAMWVNSLPSFRKKEFEIFYTHLGYLY  
IIVFVTHIGSDMFCMLINIFLFEFDYRLRTQSKRSLVARSAILPSNLLETFSK  
SLLDLESTPLSLFHPASTSIQMHGPELTSSNLEKOTLSVIRIROSQWOKLYTHLS  
DVLIDSEKHYHDLAFLDITPLDMSADISRLNRIEAYITREDKPEPTDQHTLLO  
TYKREPORLPDSIPVIGAPNANFLVGLIYSLSFVFLILGIVTRYIYIVPHNTGSL  
YNSYERILMLPGSLCAFISSVYVFLMRKONKGGKDEFKNOQVSEKOPTSPSGS  
KAKCSOLAN(LHLEAS)FN"  
complement(join(20835..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.9"  
/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
/protein\_id="BAB08726.1"  
/db\_xref="GI:97588227"  
/translation="MGNMSLVKLMVLLFLGNIFVIMISTNLFKSKMPTLSKYLY  
TTPYRQCPNTVLLTLPKMFIALVSCVYLRIRKLSQSYLYEEREMKLKIMKRVIN  
MNPGLIYATBELTSLRLVALLVMSLYNLYLYVIEGHANNDQAKIMQAFRAGELI  
GYVNCILAELEFPYTRASTLTPLVGLTSESSIKYIIMIGHVNSPCLVHTVFLIY  
AMIKMLKEFAMNPYVNLNAGTIAVIGIAMWVNSLPSFRKKEFEIFYTHLGYLY  
IIVFVTHIGSDMFCMLINIFLFEFDYRLRTQSKRSLVARSAILPSNLLETFSK  
SLLDLESTPLSLFHPASTSIQMHGPELTSSNLEKOTLSVIRIROSQWOKLYTHLS  
DVLIDSEKHYHDLAFLDITPLDMSADISRLNRIEAYITREDKPEPTDQHTLLO  
TYKREPORLPDSIPVIGAPNANFLVGLIYSLSFVFLILGIVTRYIYIVPHNTGSL  
YNSYERILMLPGSLCAFISSVYVFLMRKONKGGKDEFKNOQVSEKOPTSPSGS  
KAKCSOLAN(LHLEAS)FN"  
complement(join(20835..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532





gene  
CDS  
AAAGTCEKIGEREPEVLMKKPLD"  
/complement(3568..4629)  
/gene="psbA"  
/complement(3568..4629)  
/gene="psbA"  
/codon\_start=1  
/product="D1 reaction center protein of photosystem II"  
/protein\_id="AAFA3791.1"  
/db\_xref="GI:7259490"  
/translation="MTATLERRESANLMGRCFETSTENRLTYIGMFGVIMIPCLITTA  
ISVYIIAFAVPVDIDGIREPYSGSLGNNITSGSVIPMSNATGLAFYPIWEASL  
DEMLYNGPGLAVVCHPLGTLACTYMGREMLSPRLMPMAVAYASAPAAATAVFLI  
YPIGGSFSDGMLPLGISTENMIVFQAEHNIIMHPHMGVAGVFGSGLEFSAMHSL  
VTSLLIRETENSANAGYKFGQEEETNTVAAGHFGPLIROYASFNRSLSHFEFLA  
VMPVYIGMTAMGISTMAFNMGFNFNQSVSDSGRVINTWADIIRANLGMVMEHR  
NAHNEPDLASVEAPAVNG"  
/complement(4754..6550)  
/gene="fts1"  
/complement(4754..6550)  
/gene="fts1"  
/codon\_start=1  
/product="putative peptidoglycan synthetase"  
/protein\_id="AAFA3792.1"  
/db\_xref="GI:7259491"  
/translation="MNKKRIGFSRIVLWILFFSGSLLGLRFLYQVMKSGMLTK  
AKNQITILNTPQPRRTICDRNGIPLAIDTAYDIFAHPLHFKRSTEDIANELSLIM  
LDVEYLONLFINNTGICILAHAPROANOILAKNIEGIELEHOPKRYRYPKCADY  
IGVNTLHEGQAGELSCQESLQDLPREYVGAIDGRGLINDGIPRELFKODSCLQL  
ITDLDOKASVLAIDGIRKCAKNGIYIILDYPCALALVATAPSYDPNYYDPE  
RKNMPVIDLYEPGSTFKPINMAIALKAKIKNDPFYDQICGLGNVFLPTEPPLS  
OPACKNSHLINTDYSNSNGVNHILQRLAPELYYQIQICGLGNVFLPTEPPLS  
SYSLSKLTLEFTSYNESAVTSFGQGLAMTPKRLQVLACLSNGCNTIRPYVQGLFD  
IONERKLPFNNIPONISLKRKLKTKVSPSTETIYLDMLEVIFNGTSGSCITPG  
YRLGKTGTSOKHAROGYSTKHLITLSPFAIRPINNPQVYLVIDEIPLSFSNT  
ACPVVRSILSLIRKIKPSPITLHHYCK"  
/complement(6788..7477)  
/gene="clpP"  
/complement(6788..7477)  
/gene="clpP"  
/codon\_start=1  
/product="proteolytic subunit 2 of clp protease"  
/protein\_id="AAFA3793.1"  
/db\_xref="GI:7259492"  
/translation="MPGIPKVAARIPEAVSTYDVNRYRLYERILFLGEDLDEVA  
NOLIGVAVFLNSDDTGITFYINSBGSGMAGIYVDIMQHNVDVTTICMGLASM  
ASFLAGCTGQGLMPHARVMDHOPMGCGAKYVWESEVYKRLRELIALYLYKR  
TGSLEKIRKDMRDNMRPQAKKEYGLIDHVTNVNDELDELQKSDMLKIKVNLIT  
NIETSNKSEK"  
7825..9351  
/gene="psbB"  
7825..9351  
/gene="psbB"  
/codon\_start=1  
/product="CP47 chlorophyll a apoprotein of photosystem II"  
/protein\_id="AAFA3794.1"  
/db\_xref="GI:7259493"  
/translation="MGLPMYRVHTVVLNDPQRLISVHIMTGLVSGNAGSAMFVFLAV  
FPPSPVLNPMWROGMFVLPENTRLGISKSWGCMINDSTIDELMSTBEVAAHTIT  
LAGLFLASMMHWYWDLELFRDPRTGKPADLPIFGIHLFLSGLLCFGFGAEHVTG  
LFGPGIWSDPYGITGRQPIPEPSWAGDFDPFNGIASHHIAIGIIGLAGLEHLS  
VRSERFLYKALMNGVEYVLSSTIAVMAFVAGSTGMWYGAATPIELFGTRQWD  
LGIFNKEINKRQVASIASGSTASEMSRIRPELAVYDYGNNPARKGFLRAGAMNGD  
GIAAGNLGAHVAEKDGRGLFVRMRTFFETFPVLLDKDGIADIRADIPRRRESYSI  
EFGVSAVAYGGLDGVTPKDEPTVYKVARAOLGIEFEDRARKLSCGVPRSSRGW  
FTTGHCLFALLFFGHIMWGAFTIRFDVPAIGIDPDLEQVEFGAFQKLDASTKQAV

gene  
CDS  
/db\_xref="GI:7259494"  
/translation="MEALVYTFPLVNGTLGIIFFAIFRPPRIIK"  
/complement(9669..9800)  
/gene="psbN"  
/complement(9669..9800)  
/gene="psbN"  
/codon\_start=1  
/product="N protein of photosystem II"  
/protein\_id="AAFA3796.1"  
/db\_xref="GI:7259495"  
/translation="METATFISFSCLLIGLTGYSLYTSFGNASSLRDPFEHED"  
/gene="psbH"  
9891..10112  
/gene="psbH"  
9891..10112  
/gene="psbH"  
/codon\_start=1  
/product="10 kDa phosphoprotein of photosystem II"  
/protein\_id="AAFA3797.1"  
/db\_xref="GI:7259496"  
/translation="MADTSQGRYVGNFLKPLNSEYGVKVAAGMGTIVLMGFALFA  
VFLVITLIQYNASVLLDGIIPANWSSLKY"  
10445..11092  
/gene="psbB"  
10445..11092  
/gene="psbB"  
/codon\_start=1  
/product="apocytochrome b6 of cytochrome b6/f complex"  
/protein\_id="AAFA3798.1"  
/db\_xref="GI:7259497"  
/translation="MSKRYDMFDRLEIGIADITSKYVPPVNIYFCIGITPFC  
IMVAGSFAMTFYRVPTEAFASVOYIMDVNFGMLISIRKMSMAVLMILHVF  
RYLTGCFKKPRELTVTGILAVCTVSEFVGYSLPMDVGWAVKIVTGPDAIV  
IGAVLVELLGGVGVGOSTLRFYSLHTEFVLPPLTAVFMIAFLMKIKGIDGSLP  
11202..11684  
/gene="petD"  
11202..11684  
/gene="petD"  
/codon\_start=1  
/product="subunit IV of cytochrome b6/f complex"  
/protein\_id="AAFA3799.1"  
/db\_xref="GI:7259498"  
/translation="MSLOKKRDLKDPVLRALAKGMGHNYGEPAMNDLLYTFPVC  
LGTICGLAVLEPTMGCEPAPNPAPLELIPMPFPVQIIRVIRPNKILGVNLAA  
GVPAHLITVPIESINKFQNPFRRVAMTVLLIGVAVVWMLGIGATLPIIDRSLMGFF  
complement(11779..12192)  
/gene="rps9"  
complement(11779..12192)  
/gene="rps9"  
/codon\_start=1  
/product="ribosomal protein S9"  
/protein\_id="AAFA3800.1"  
/db\_xref="GI:7259499"  
/translation="MSVOKMPPILNYLGTGRKSAVARYVPGNGEVLINGLPCTNY  
LOFNGFLSAVRSPLLETLDYEDYDIIVAVAGGLTGQAEAIRLGVAAALCTIDISNR  
HPLKGGFTLRDSRVKRRKYGLKAKKAPDPSKR"  
complement(12245..13195)  
/gene="rpoA"  
complement(12245..13195)  
/gene="rpoA"  
/codon\_start=1  
/product="alpha subunit of RNA polymerase"  
/protein\_id="AAFA3801.1"  
/db\_xref="GI:7259500"  
/translation="MSGNDLFPSTIYCTIESKIESPRNLYGRFLIEPLAIGGITVGN  
LRLILGIESAITSYKIRGANNEFSILFGIRSEVLEILNLKEIVFRRKSLDVQK  
YLSIOGCVKAAANLQPLTSIEVDGGQYIATLSGNMLDMEFVINYGKGYOMADYAI

Query Match 17.4%; Score 205.8; DB 6; Length 118360;  
Best Local Similarity 60.3%; Pred. No. 9.3e-37;  
Matches 363; Conservative 0; Mismatches 237; Indels 15; Gaps 2;  
QY 229 agtcgcgaatcgcagccgacgttggtctacgtacccgtctctctcgtcgcga 288





Db 98658 GGCATTGCTTTGMAATGCGCTAGAGATTAATGCTAACA 98615

RESULT 6  
LOCUS AP001129 194509 bp DNA PLN 05-FEB-2000  
DEFINITION *Oryza sativa* genomic DNA, chromosome 6, PAC clone: P0644B06.  
ACCESSION AP001129  
VERSION AP001129.1 GI:6907081  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail: tsasaki@agr.affrc.go.jp, Fax: 81-298-38-7468) URL: http://www.dna.affrc.go.jp:82/, Tel: 81-298-38-7441.  
The orientation of the sequence is from T7 to SP6 of the PAC clone. Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against the non-redundant database NRP (PIR, SWISSPROT, RGP, Protein similarities of the coding regions were searched against NRP with BLASTP2.0. ESTs represent the identified cDNA no. and RCP clone ID.  
This sequence of this clone has an overlap with P051612 clone, DDBJ: P0000616 at the 3' end. This clone ends at the position 30,063 of P051612. Detailed information on overlap and assembly quality together with annotation of this entry at <http://www.dna.affrc.go.jp:82/genomicdata/genomefinished.html>.

## FEATURES

Location/Qualifiers

1..194509  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="6"  
/clone="P0644B06"  
/join(53..107,495..542,2129..2223)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA90609.1"  
/db\_xref="GI:6907082"  
/translation="MHACOYCYKRLITLVYVPHIVTSKVLIMHAACSSYLARYLQV QIPRMLOKHVOOHESSAKV"  
complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))  
/note="ESTs AU082304(C61278),C93463(C50982) correspond to a region of the predicted gene.; Similar to Arabidopsis thaliana DNA chromosome 4, BAC clone F24A6, Pollen-specific protein precursor like. (AL035396)"  
/codon\_start=1  
/protein\_id="BAA90610.1"  
/db\_xref="GI:6907083"

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))

## CDS

complement(join(6634..6742,6856..7058,7301..7379,7465..7830,8543..8909,9188..9326,9414..9684,9802..9911,10520..10657))













- AUTHORS Douglas, S.E. and Durnford, D.G.  
 TITLE Nucleotide sequence of the genes for ribosomal protein S4 and tRNA(Arg) from the chlorophyll c-containing alga *Cryptomonas phi*  
 JOURNAL Nucleic Acids Res. 18 (7), 1903 (1990)  
 MEDLINE 90245597  
 REFERENCE 4 (bases 18535 to 19351)  
 AUTHORS Douglas, S.E. and Durnford, D.G.  
 TITLE Sequence analysis of the plastid rDNA spacer region of the chlorophyll c-containing alga *Cryptomonas phi*  
 JOURNAL DNA Seq. 1 (1), 55-62 (1990)  
 MEDLINE 92119320  
 REFERENCE 5 (bases 34539 to 35380)  
 AUTHORS Reith, M. and Douglas, S.  
 TITLE Localization of beta-phycoerythrin to the thylakoid lumen of *Cryptomonas phi* does not involve a signal peptide  
 JOURNAL Plant Mol. Biol. 15 (4), 585-592 (1990)  
 MEDLINE 91338697  
 REFERENCE 6 (bases 110917 to 113854)  
 AUTHORS Douglas, S.E.  
 TITLE Unusual organization of a ribosomal protein operon in the plastid genome of *Cryptomonas phi*: evolutionary considerations  
 JOURNAL Curr. Genet. 19 (4), 289-294 (1991)  
 MEDLINE 91330343  
 REFERENCE 7 (bases 40675 to 42376)  
 AUTHORS Douglas, S.E. and Turner, S.  
 TITLE Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor  
 JOURNAL J. Mol. Evol. 33 (3), 267-273 (1991)  
 MEDLINE 92099311  
 REFERENCE 8 (bases 96129 to 98906)  
 AUTHORS Wang, S.L. and Liu, X.Q.  
 TITLE The plastid genome of *Cryptomonas phi* encodes an hsp70-like protein, a histone-like protein, and an acyl carrier protein  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991)  
 MEDLINE 92073372  
 REFERENCE 9 (bases 106789 to 108216)  
 AUTHORS Douglas, S.E.  
 TITLE A secY homologue is found in the plastid genome of *Cryptomonas phi*  
 JOURNAL FEBS Lett. 298 (1), 93-96 (1992)  
 MEDLINE 92183838  
 REFERENCE 10 (bases 42198 to 44153)  
 AUTHORS Douglas, S.E. and Reith, M.E.  
 TITLE A *hchl* homolog, encoding a subunit of Mg chelatase, is located on the plastid genome of red and *Cryptomonas phi*  
 JOURNAL J. Mar. Biotechnol. 1, 135-141 (1993)  
 REFERENCE 11 (bases 82327 to 84479)  
 AUTHORS Douglas, S.E. and Murphy, C.A.  
 TITLE Structural, transcriptional and phylogenetic analyses of the *atpB* gene cluster from the plastid of *Cryptomonas phi* (Cryptophyceae)  
 JOURNAL J. Phycol. 30, 329-340 (1994)  
 MEDLINE 912 (bases 98901 to 114602)  
 REFERENCE 12 (bases 98901 to 114602)  
 AUTHORS Wang, S.L., Liu, X.Q. and Douglas, S.E.  
 TITLE The large ribosomal protein gene cluster of a cryptomonad plastid: gene organization, sequence and evolutionary implications  
 JOURNAL Biochem. Mol. Biol. Int. 41 (5), 1035-1044 (1997)  
 MEDLINE 97283757  
 REFERENCE 13 (bases 61067 to 68605)  
 AUTHORS Leitsch, C.E.W., Kowalik, K.V. and Douglas, S.E.  
 TITLE The *atpA* gene cluster of a cryptomonad, *Galliardia theta*: A piece in the puzzle of chloroplast genome development  
 JOURNAL J. Phycol. (1998) In press  
 REFERENCE 14 (bases 1 to 121524)  
 AUTHORS Douglas, S.E. and Penny, S.L.  
 TITLE The plastid genome of the cryptophyte alga, *Galliardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae  
 JOURNAL J. Mol. Evol. 48 (2), 236-244 (1999)  
 MEDLINE 99128221  
 REFERENCE 15 (bases 1 to 121524)  
 AUTHORS Douglas, S.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JAN-1998) Institute for Marine Biosciences, National Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1,

COMMENT On Sep 15, 1998 this sequence version replaced gi:11396 gi:11297  
 gi:18103 gi:18281 gi:11383 gi:11407 gi:12539 gi:336730 gi:11300  
 gi:398949 gi:2661180.  
 FEATURES  
 source Canada  
 location/Qualifiers  
 1. 121524  
 /organism="Galliardia theta"  
 /organism="plastid:chloroplast"  
 /db\_xref="taxon:55529"  
 /note="plastid"  
 complement(5..86)  
 /gene="trnY(gua)"  
 /note="codon recognized: CAU"  
 /product="tRNA-Tyr"  
 /anticodon="(pos:50..52,aa:Tyr)"  
 complement(5..86)  
 /gene="trnY(gua)"  
 complement(181..573)  
 /gene="rpl19"  
 complement(181..573)  
 /gene="rpl19"  
 /codon\_start=1  
 /product="ribosomal protein l19"  
 /protein\_id="AAC35594.1"  
 /db\_xref="GI:3602933"  
 /translation="MEIILHLEFVNMKSNKNILOLVESNHSNKTLLPOLQGVTVRLG  
 VSIENKRIQATEGVITSKNSGIMLTITIRRVLGIGIERVYLHSPKLSIELT  
 RSKIRRAKLYLTRSGKATRILOQRN"  
 729..3188  
 /gene="clpC"  
 729..3188  
 /gene="clpC"  
 /codon\_start=1  
 /product="Clp protease ATP binding subunit"  
 /protein\_id="AAC35595.1"  
 /db\_xref="GI:3602934"  
 /translation="MEERTEKAIVIMLAQEARLRGNFVTEQIILGIGTGTI  
 AKVLSKMNKLDARVEVEKILGRSGSFVVEIPFTIRARVLELSSEEROLGHW  
 TGTTHLLGLIREGEVAAVLENLALDTKRVYIILDDFAVLSATKQGTGTP  
 TIEFGSLNLTOKAAGKLDPEYVROKTEIRVIOIGRTKNNPILIGPGKTAIAL  
 GLAQRNNNDVDPDILIEDKRVVPLDGLLVAGVKKYGEFEERLKKIIDIRVANNVILV  
 IDEVHTLIGAGAEGALIDANTIKKALRAGKDEMOICGATLEERKIKDSLEKRO  
 PWWGEPSVEETIETILGRDRYEHKHKYISDEALSAAKFADYIDRLPDAID  
 LIDEGSRVRLNNSQDLPAPARELDELRLEITLQKDEAVSODFETAGOLRDEMETIA  
 OLAIHAKSKGDEENKREVSVTEEDIAIVAMGIVPNKMTRESKILQMETEIKH  
 GRITIGDEAVVAASKAIRARVGLKIGSPGVYNGGQLVESVRRPITYVLPDIKCH  
 EEMVRLDSEVEMERTVSKLIGSPGVYNGGQLVESVRRPITYVLPDIKCH  
 PVPYFNILOITLEDGRLLDQSGRTVDKNTLLILTSNVGSKVLEKGGGLFDSQI  
 ESOYGRKALVNEIKOYRPEPRLNLDIIFRQITKDEVEIEIMLKEVFTRISE  
 KGIOLETVARFKHLLINEGYNPIYGARVPLRAVMRLIEDTLSEEFIAEKIKGDAV  
 DVDDGKRVLLGSKLELVAN"  
 3316..3388  
 /gene="trnF(gaa)"  
 /note="codon recognized: CUU"  
 /product="tRNA-Phe"  
 /anticodon="(pos:3349..3351,aa:Phe)"  
 3316..3388  
 /gene="trnF(gaa)"  
 complement(3402..5201)  
 /gene="dnab"  
 complement(3402..5201)  
 /gene="dnab"  
 /codon\_start=1  
 /transl\_except="(pos:complement(5199..5201),aa:Met)"  
 /product="replication helicase subunit"  
 /protein\_id="AAC35596.1"  
 /db\_xref="GI:3602935"  
 /translation="MKVTRFYNLAKAEKHLIIQILVEKDFEIQITKIDTQVYFROSHO  
 ILYRSINILHTRKVLINLPNLISLHSELVSEVYINVTITLFNDNRFLININDSK  
 IIVDNFIRELONSCAKIVSLTFNLSLSEVTLQSKNSLILNINYNKSKLSKISOL  
 LLEETLEIDKTNRESHVLTGPFDDHLLVGLQSDLLIIGASPMCKTAAEMLSLVN  
 VADIOSFVIVISLEMSKQILVRLISMETNATSRLEGNISISEWILRAMTILS  
 NNTIYLDENNDIVDIOKSLASLOQIIGDGLIYIDVLQLOKREKSSQRIHLSYI



gene  
CDS  
/gene="ppk"  
/complement(1489..3675)  
/gene="ppk"  
/note="ORF\_ID:s110290"  
/codon\_start=1  
/transl\_table=11  
/product="polyposphate kinase"  
/protein\_id="BA10660.1"  
/db\_xref="GI:1001781"  
/translation="MIOAMPAPTVDFKDPAYFNNELSMALFNQVHLHGLDRTPL  
LEKLFALFCSNLDPEFVRVAGLKOQVANYTKLADRTSPQOLKEISLRPLV  
YQONVFEVYLFKFLADEGIFLNDYDLSQERQYHOFDDHIFPVLPLADPSHP  
FPIYSNLSINLGVVADPDDELFRKVPPLPRVALPDVCOPEPKPMLMTG  
PLEQVLAHNLASLEFGMIOEHLFVTRNADVADEEDDLIAIEPLRRRKG  
SAVLEINASTPKNIDRLMTDLGLEIDYDIDGLDKLFFLSLAPHLKDEPW  
ASVIFPRKHYEFVDEDEGDEGVQOEGIDIFTLIRQDILVHHYOSFTASVOOFITQ  
AAYDPHATIKMTLRTSGDSPYVNALIAAENGKQVAVLEKARPEENINIMAK  
LEQYCHAVYGIVGDKTKHTKTVLVROEGDIRRYVHIGTGNPKAKLYTDGLIT  
CRELGNDLNLNLFELTGSROQDYOGLVAVPMREPMAMERADHNGSTGRI  
VAKNLSLVDTQIIIRALYASQAGYQIDLIVRGICCLARGVENSENIRVLSIRLE  
HSRIFFHNGEIEIYIGSADWMSRNLFRVNAVVPYEQDIOEQIISIGILLADR  
QAMEIAPDGTIVORRPPASPEOSQSAIFTAQALIAETTEDELRSV"  
/complement(4315..4608)  
/gene="minE"  
/complement(4315..4608)  
/gene="minE"  
/note="ORF\_ID:s10546"  
/codon\_start=1  
/transl\_table=11  
/product="septum site-determining protein MinE"  
/protein\_id="BA10661.1"  
/db\_xref="GI:1001782"  
/translation="MILELIRFSRSGKNGSGEDARRRLKLVIANDRSGLSPEMEEM  
RREIVVSRIVEIDPGMEESLESQDMTALIANLPVRVRRRTAKSQAOS"  
/complement(4666..5466)  
/gene="minD"  
/complement(4666..5466)  
/gene="minD"  
/note="ORF\_ID:s110289"  
/codon\_start=1  
/transl\_table=11  
/product="septum site-determining protein MinD"  
/protein\_id="BA10662.1"  
/db\_xref="GI:1001783"  
/translation="MNRIVTVSGKGVGRTTTANLGAALRLGKVVYLIDAEGR  
NIDLLEGEORIVYADIVADECTIDKALVKKRLPNVLPAQNRKSDAINAKOM  
QSLVLEKDKPXYIITDCPAGIEAGFRNALVAPQEAIVTTPMSAVRADRYIGLE  
ADEIGKISLIVNRILPREWOLINOMISVEDLIDLAPVIGILPDQKIIISTNKGRL  
VVEEKLSPGLAFONTARLEQDIFLDPMAHNTLLRIRRLGG"  
/complement(5543..6001)  
/gene="minC"  
/complement(5543..6001)  
/gene="minC"  
/note="ORF\_ID:s110288"  
/codon\_start=1  
/transl\_table=11  
/product="septum site-determining protein MinC"  
/protein\_id="BA10663.1"  
/db\_xref="GI:1001784"  
/translation="METNRQTAVAASAGLSVDTQADKPLVDPSETPSLPLPLVVR  
HTRSRGEIRHGDVYITIGVNGSSIVAGDILIMCCLGMHAGKQDQAVIMLL  
RLAACQIRICDRALARVADAVDRREPILVITSGILTFVQRFQSFALL"  
/complement(6028..6285)  
/note="ORF\_ID:s10543"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA10664.1"

gene  
CDS  
/db\_xref="GI:1673351"  
/translation="MSDTPHEWOLVPOGNFVRLSTLPVTATPDSPLVESHQQL  
EOSTGNVIRIAGRWODOTLVLAANNQDLADRQALATALA"  
/gene="frah"  
/note="ORF\_ID:s10298"  
/codon\_start=1  
/transl\_table=11  
/product="heterocyst to vegetative cell connection  
protein"  
/protein\_id="BA10665.1"  
/translation="MAVCEPCNHTNPNAGIQCACPTPLPTLMPCFGSSVQSDATF  
CGNCGOTIPPOGLVSGAPSEEEVPIIAPVESAEPPIAPVSTPAPVQTOLOS  
OGASIVHONQOTLEIPAGITVHLKPNKVPPIIDVSGFANSVSRHADIRLEG  
NEYFLEDVGSANGTYNHTLLFEGNHLLARPGRLALGKGLVTFIFQLH"  
/note="ORF\_ID:s1r0299"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA10666.1"  
/translation="METPVKQPRLLPPEERAGYLDMSDFETDRLQIPYIDLTLQ  
LDVYNAGVYQOQKDSQSGFFALILHMMQCLQNHMPRIRYNQWYVLDNPPAMYS  
VALGKERFSEMLLENVQSLSYGFADYISKLQITAGNEREYVPLACIAGNL  
PALKFTLSVHYRREKIQOGWFFYGGRYWODKLLIPLCKIRHANADPVLQDLIS  
QFORLFSMPCIPSSS"  
/note="ORF\_ID:s1r0300"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA10667.1"  
/translation="MYSTILKRLPIISLDOQFVALCOQNPDLKFRSSNGDLITMP  
TGGPFRNSSELNDFNLNRRTOGVFDSSTCFRLPNGALRSPDVAMVRSWEAL  
TPAREKRPPLCPVLELSPSDALATQAKMREYLSCGTELGLINPIDDQVEIYR  
PDSPLEILAKPOTLNGDRLPLDELNVAMKCAN"  
/note="ORF\_ID:s1r0301"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoenolpyruvate synthase"  
/protein\_id="BA10668.1"  
/db\_xref="GI:1001788"  
/translation="MVSVAVERKSVAKHETALILMEEVGTHDVLGKNSLGM  
OOLTNKGVNVPSGFATTAAYARFIEQAGLEOKLDFLDVNDMANLQERHLLAKQ  
LITDPPEPNOATAIEAVGAMGERYQKGRGVAVARSSATADLEPASVAGQE  
TILNHSLSVLESCRCFRSLFTDRAISRHNGNDDEYVETKLEKGRPLIEKRIG  
GVNESIDTETGERKNAALITTAAYGLGENVVAGANPEYEFKTEKLEKGRPLIEKRIG  
SKAKMYVDVGSRLKTRNVEVAPREKKTINDELTOLARACIIIDHSGVAGVYT  
PMDIEWAKDQOTGELFIVARPEFTVQSASANYTKYTELAKDSQVATGRVGAALGO  
GKAQVIRNYSQINQPRGVELITNRDPPMEIKMOKASAVITNQGKTKMAAIIAREM  
GIPAIYVCGDADDTIKTGEDVYITCCSGDEGSYSGILNVEVHETELSNLPRTQTL  
MNYGNPEOAGFASYPADGVGLARLEFIINHHKAPLALMKFDELEDPALAKAIAL  
TKLYAGDRPEFYDKLAHGIAMIAAAYPKVYVVRMSDFKSNYANLIGROFEPKKE  
NPMIGWGAARYYDPNTRKALVLEQALKVROEMGLTNIPIPIFCRTPEDEGKAVIA

Query Match 10.2%; Score 120; DB 29; Length 135638;

Best Local Similarity 51.0%; Pred. No. 3.8e-17;

Matches 347; Conservative 0; Mismatches 315; Indels 18; Gaps 2;

Qy 193 caatcgcaaccagagctcgccgagacaccccaagatcgctgcaaccgagccgaact 252

Db 5404 CAATTAGAGGCGCCCTAGCTGATTGGGAAAAAGGTTGTCCTGATTTGAAGCGGACTT 5345

Qy 253 tggctacgtaccctcgtctctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 312

[illegible]

```

/product="reverse gyrase"
/protein_id="AAC07000.1"
/db_xref="GI:2983417"
/translation="MikAIPDTPKCGGGEISAEKRLKGLPCCKLPPEVNRREVCOK
LENGSFECELLSEKLDWDFEKLCTSPMSLOKMAKRVLEGRSFMALPTGK
TFGLSMASVIAOGKRSYIILPQLVBEVSRITPCVDEDRLLVMCKLSKKKE
LKERIOGDDELLITSMFLYKNVEILPKDFSTFPDDVNSFLPTAKNDKVLTLGE
SEEDIHKAFELIRLKEKPNKSEEDMEIKRSBELRLAKKKGVLAASANGNSGR
RIKLEFRELGEVCKPSVLLRNIVETYEETONKETLTKRVKCKGGIIVFSDGR
BAVEKPKLESHGVKAVTEEDLKEKEGAOYAKISKFNPLAGIDLPVYARA
IFVGPRIKPLKYETSVSMLLALSLRPIILKDEKSEVKVDWIQRLRISFL
SDVEERKPLKDRITENLRKVOBELREDIVEIKNSEELTLRAGEGEVAVADT
GYLQASGRTSRMAGGLKGLSHVINDRRAFKMLKKVRMENDINPKIEMDKLE
VLREIDERRKRVIRLEGRVAKOEKHNKPYLVVESPNKARTIANFGKPMGKICG
IDVLEVMAGDLYIMTASLGHVFLDKKEPHYIAKNGEVPIYEVIEGENYKGL
RELAEVDVTLIGDPPDEKIGMDLGLALLSPYIPANERIEPEHEVRKAKIENVP
RDENILYKAOVRIADRWGFEVSRVQOAFKHLMSLGSROVPIUGMTIEREKY
RKKHVNQITRENGRMRLGFEFQDKKAEKPEYNLEIIVLEVEEBELKEMNISEE
TJDTMLKASREGGAGHNCIRPTKPLVEDLRSMLAGOLNLRHEHLLALELIPK
LKELEFPRVMSGEGAKVAKALGDEOELLITTELEGFNKVYPLELOPDLKSVYEDK
MASOKPVAKVTKKVKVAKALGDEOELLITTELEGFNKVYPLELOPDLKSVYEDK
KELSVPAVLYTQGSVLEEMKRGITGRSPATVTSKILLEGYVIERHGLIPIKLG
KQVEFLKREKIMPEVSEETRLLELMDKVEEGEDYLOVLDLYKKVNEFEKANV"

```

```

gene
8016..8462
/gene="aq_888"
8016..8462
/gene="aq_888"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07002.1"
/db_xref="GI:2983419"
/translation="MGTMLFDEMYQKGOIIVHEVSKELGIDEHKAFLIRAVIOTL
RDRPASEGKDFPAOLPMVLKAVMCDMDPTRPDSIKRKDFLEVMNHPELRPA
DIESLEDERVYAVFVILAKRHISYGEIKVYSELPELREMLENA"
8506..9309
/gene="aq_890"
8506..9309
/gene="aq_890"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC07005.1"
/db_xref="GI:2983422"
/translation="WRVREPAVAGTFYFKDKBELKMLDLGCEPEKRIKPAIIVP
HAGYIYSGTACEYKRIETPEKAVILLPNHGTGKPISVSGDAMEPYGVLEIDG
LREKILKYPVANDPEYAHLYSHLSVQLPFIQRYARRFEKILPIVITVEVAKDFG
RFLGELIKEDALIVISSDMHYVPAEERKDELLISAMERINTEELYAVAVQYNT
MCGVPAVVGIESAKVLCATKGIIVVDSGSDTGDYISQVAYILGMITL"
complement(9289..10227)
/gene="pax"
/note="aq_891"
complement(9289..10227)
/gene="pax"
/codon_start=1
/transl_table=11
/product="exopolysphatase"
/protein_id="AAC06997.1"
/db_xref="GI:2983414"
/translation="MSLDNKPIMRVASIDIGSYSVRLTIAQIKDKGLSILLEGRTS
LTKVAKETGLOEDRIEFTIOVLKEKKLIDEFKVEKVAATAEALRRANAEFLER
VGRVGLVEVITIPBOGRAYILAVAYSLKPEGVCSVDGSGSTEVEFGKRVRE
ISLPIGIVNLTEFFKODPTPEEKRFEPFELKELSKVKPVDITVGLGGLITTLAA
LEYWVYDPQKHGKVLTYGQIKKMDITKEIPSEPRSKRFQVEDRRAKVILAGIG
TEKLTLEFEKDLIVSDWGLREGVILSEVLEKENS"
10187..11167
/gene="fabd"
/note="aq_892"
10187..11167
/gene="fabd"
/codon_start=1
/transl_table=11

```

```

/product="malonyl-CoA:Acyl carrier protein transacylase"
/protein_id="AAC06999.1"
/db_xref="GI:2983416"
/translation="WDATLITGLSKDMKRIAYVPPGSGSYGMGEFYKOFHEIAD
VYHAEHLRMSLDDIDVMEKGRPEKRLKTVYQPMALSLIYALESMGPKPD
PVAGHLEIYALAVAGIDVOLAERKLYNKAQIMOEAPVPCAKMAVAILKPEKVE
FACERAKDGVVEPANYNSSTQVTSIGSRPAVEKAAQIAKMGCRALPIKYSVSHS
LMKPAADPRLKLAOAPIONIKIPLVQNYTKAHTMAHEIRNNLHOIFSVKMYQSV
TMHMQGDTPEIIPKMYLSKLIQOTLPNVFVNEKPEDAESVLKAT"
11164..12015
/gene="aq_893"
11164..12015
/gene="aq_893"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC07006.1"
/db_xref="GI:2983423"
/translation="MRAHPCIPCLINOGINAVKRLNLPREKEMETATRSKLFLOPD
TPRSPAYVAVFIOQIVKEISGEDEPKDLKLANEAGBELSKVLKGPDLKALKL
SAIGNFIDFAIKVDFERLOKLPIDYDKFLERLENAKSVFIIGNAGCIV
FDKVLVETLKGMEKVVYGVKRPILNDATLEDAEVSMTKLCVIDNSDKVYLE
DKKEFEYETFSYSDYISKQANFETLSADRDLFLVAKCDPIGIREFGKRGELIF
KYRPSKD"
complement(12004..13008)
/gene="quea"
/note="aq_894"

```

```

Query Match      8.1%; Score 95.4; DB 11; Length 14103;
Best Local Similarity 52.5%; Pred. No. 1.5e-11;
Matches 260; Conservative 0; Mismatches 226; Indels 9; Gaps 2;

QY 229 agtcgtcgcaatcgacgcgcgtgtgtctacgaacctgcctctctcgtcgtcga 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 840 AGTCCTTTAATGAGCAGACATCGGTTGAAACCTCGATGATGATGATGATGATG 781
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 289 aaacgcgtcaataacacgcgtcgtgtaagtttcaacgcgcgtatggaacgaac 348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 780 AAMCAGGATGATACGACATCTTGTGACGCTTGAAGGAGGCTCCTTATGAAAAGC 721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 349 cctagtcgtgataaagcgtgcgaattcgaattcgttattcaaacctagtc 408
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 720 ACTGTTAGAGACNAA--AGGGACTTTCCTATGCTATGCTATGCTATGCTATGCT 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 409 aaattcccttagatttgggggaaagctttagtttgcgttagatcataaagatag 468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 663 AAATAAGACGTAATAGACATTGAAAGTGAGACAAACTGTAGAACAGATTAAATTC 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 469 gcaagaaggtgccggaatttatactatagattgtccgtcagattatgcgggt 528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 603 -----AGGAATTACGATTACATCTCTGTGATTCTCTGCTGTATGAAAAAGTTT 550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 529 cataacgcattacacgcgcgttaacgaacgcgtatagttacaacacctgatlactgc 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 549 TCAGATAGCGGTATCCCTCGCGACAAAGCTTAATAGTTTATCCAGAGTTTCCTC 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 589 attgagagatgcagatagatgattacaggtcttgatgattgattgattgattgatt 648
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 489 TATAAGGACGCGAGACAGGTAATAGGCTTCTGACAGCATGACAGAGAAATTACAA 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 649 aatgatttgaacagagttgaacagatttgaaggggtgaagatgattgattgatt 708
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 429 AGTTATCTTAACAGGATAAGTGGAATGTGTAAGGAGGCGCTATCTTACGATTGA 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

QY 709 tgatgttcaagagat 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 369 GGATATAGTGAATAT 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 13
AF149810 673 bp DNA PLN 01-OCT-1999
LOCUS AF149810/c
DEFINITION Oryza sativa cell division inhibitor Mind homolog gene, partial

```

ACCESSION	AF149810
VERSION	AF149810.1
KEYWORDS	GI:6002787
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Magnoliopsida; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE	1 (bases 1 to 673) Han,F., Kilian,A., Chen,J.P., Kudrna,D., Steffenson,B., Yamanoto,K., Matsumoto,T., Sasaki,T. and Kleinjofrs,A. Sequence analysis of a rice BAC covering the syntenous barley Rpg1 region Genome (1999) In press 2 (bases 1 to 673)
JOURNAL	Han,F., Kilian,A., Chen,J.P., Kudrna,D., Steffenson,B., Yamanoto,K., Matsumoto,T., Sasaki,T. and Kleinjofrs,A. Direct Submission Submitted (12-MAY-1999) Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA
TITLE	location/Qualifiers
FEATURES	1..673
source	/organism="Oryza sativa" /cultivar="IR-BB21" /db_xref="taxon:4530" /chromosome="6" /note="covering the syntenous barley Rpg1 region" complement(<1..537) /product="cell division inhibitor Mind homolog" /protein_id="AAE00142.1" /db_xref="GI:6002788" /translation="MAFAPRLLLSRCPPSPARRHGRTAPELGGPRVWVTSG KGVGKTTTNTAASLARLSAVADADGLNLDLGLERNVHLTAADVLAGDC RIDDLIVRHRLHQLCLSLPSKSKLPFLAFSGKTTLTVADALRRANPFAFLIDCP AGQSNAPRPSTVDGCYM"
CDS	115 a 160 c 295 g 103 t
BASE COUNT	
ORIGIN	
Query Match	8.0%; Score 94.8; DB 6; Length 673;
. Best Local Similarity	55.9%; Pred. No. 1.8e-11;
Matches 180; Conservative	0; Mismatches 142; Indels 0; Gaps 0;
Oy	193 caatcgcaaacagagctcgcgggagacaccacaagagtctgtgaatcgaagccgaagt 252
Db	11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
373 CAATCGCAGCCGCCCTCGCGACGCGCTTCCTCTCCGCCGCGCGCGCGCGCACGC 314	
Oy	253 tggcttaagtaacccttcgatttcttctcttggtcttcgaaaaacgggtcaattacacctgatc 312
Db	11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
313 CGCGCTCGCGCAATTCGACACTTCCTCCTCGGCTTGAGAACCGCTCCACGTCAACGCGCGC 254	
Oy	313 tgaagtctcaaggagcatltagactcagaacccaagccctagttcgtatataaagctgctc 372
Db	11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
253 GCAGCTGCTCGCTGGGGACTGCGCCTCGACACAAGCCCTGTCGGCACGCGCGCTCCA 194	
Oy	373 aaattcgaattgcttcttatccaacacctaaggtccaatttcgattagatttgaggg 432
Db	11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
193 TGACCTGCAGCTGCMTCGCCCTCCCAAGCCCGCGCTCCAACTGCCCCCTTCGCGCTC 134	
Oy	433 aaaagctttagttgggcttgaatgatataaagaatagaagcaagaggttcgccgatttat 492
Db	11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
133 CAAGACCCTACACMCGGCTGCTGACCGCGCTTCGCGCGCGCCCAACCCACCCCTTTTCAT 74	
Oy	493 acttatgatattgcttcgaagt 514
Db	11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
73 CTTCACTGACTGCGCTCGAGGT 52	
RESULT 14	
AP001517/c	294250 bp DNA
LOCUS	Bacillus halodurans genomic DNA sequenced
DEFINITION	Bacillus halodurans genomic DNA sequenced
FEATURES	1..74
source	03-AUG-2000

ACCESSION AP001517 BA0000004  
 VERSION AP001517.1 GI:10175500  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Bacillus halodurans DNA.  
 Bacillus halodurans  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 2 (sites)  
 Takami,H., Nakasone,K., Ogasawara,N., Hiram,C., Nakamura,Y.,  
 Masui,N., Fujii,F., Takaki,Y., Inoue,A. and Horikoshi,K.  
 Sequencing of three lambda clones from the genome of alkaliphilic  
 Bacillus sp. strain C-125  
 Extremophiles 3 (1), 29-34 (1999)  
 99184646  
 3 (sites)  
 Takami,H., Takaki,Y., Nakasone,K., Hiram,C., Inoue,A. and  
 Horikoshi,K.  
 Sequence analysis of a 32-kb region including the major ribosomal  
 protein gene clusters from alkaliphilic Bacillus sp. strain C-125  
 Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)  
 99205008  
 4 (sites)  
 Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.  
 Replication origin region of the chromosome of alkaliphilic  
 Bacillus halodurans C-125  
 Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)  
 99356711  
 5 (sites)  
 Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,  
 Sasaki,R., Hiram,C., Fujii,F. and Masui,N.  
 Genetic analysis of the chromosome of alkaliphilic Bacillus  
 halodurans C-125  
 Extremophiles 3 (3), 227-233 (1999)  
 99411980  
 6 (sites)  
 Takami,H.  
 Genome analysis of facultatively alkaliphilic Bacillus halodurans  
 C-125  
 (in) Extremophiles in deep-sea environments (Ed.):  
 HORIKOSHI, K. TSUJII:  
 : 249-284: Springer-Verlag (1999)  
 7 (sites)  
 Takami,H. and Horikoshi,K.  
 Reidentification of facultatively alkaliphilic Bacillus sp. C-125  
 to Bacillus halodurans  
 Biosci. Biotechnol. Biochem. 63, 943-945 (1999)  
 8 (sites)  
 Takami,H. and Horikoshi,K.  
 Analysis of the genome of an alkaliphilic Bacillus strain from an  
 industrial point of view  
 Extremophiles 4 (2), 99-108 (2000)  
 20263314  
 9 (bases 1 to 294250)  
 Takami,H., Nakasone,K. and Takaki,Y.  
 Direct Submission  
 Submitted (22-Mar-2000) to the DDBJ/EMBL/GenBank databases. Hideo  
 Takami, Japan Marine Science and Technology Center, Deep-sea  
 Microorganisms Research Group; 2-15 Matsushima, Yokosuka, Kanagawa  
 233-0061, Japan (E-mail: takami@jamstec.go.jp  
 URL: http://www.jamstec.go.jp/jamstec-e/dlo/DEEPSTAR/FResearch.html  
 Tel:81-468-67-3895, Fax:81-468-66-6364)  
 Location/Qualifiers  
 1. .294250  
 /organism="Bacillus halodurans"  
 /db\_xref="t"

```

gene
CDS
    236..823
    /gene="BH2879"
    /db_xref="GI:10175503"
    /product="spore cortex-lytic enzyme"
    /protein_id="BAB06598.1"
    /transl_table=11
    /translation="MMKRVLLPSCILITPFLFNKAASHSKHNOVSGPLVLLSBOY
    GVPMELIKRINERSSNTIYRGEOLITPAVATASRELLARIVAEAGEPPYAGKAVA
    VVNLNKHDPSEPDNVDVINEVNSGVYVSPONCRIPORPADEKRAVDENLORRG
    OGNSGLYFNPKTASPVSTROQITVIGNHIFAK"
    955..1605
    /gene="BH2880"
    955..1605
    /gene="BH2880"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB06599.1"
    /db_xref="GI:10175502"
    /translation="MRRISRTSSRTGRRATVYVPLGTSVSRKVEPPPIERTANET
    AKPSNNFVLSERYKQYKELKQVKSFEYHEKELYHTLQAGNKELLISQCTLLC
    KYNETLSTIKRLDLPQSTYMDIQAHLIOHOKELTELGLNDNGTLTNSDVERCO
    LSTSDQDTMAVLRKFSLLKQYHTISOLQVFNHONSHPYDPGCEPLKGTIEKG"
    1694..2977
    /gene="BH2881"
    1694..2977
    /gene="BH2881"
    /note="unknown conserved protein in others"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB06600.1"
    /db_xref="GI:10175503"
    /translation="MSEHDESHSDNLIDLMNLTITNDHLQLLKLGEMAAKRLMNL
    FLVGTVDMRGVGGDLVIEGDALFQSNVANYGKVKHHEPPATVWGAEN
    LKDIIVSRAESYAKPGALPTIRHSHITDRLARDFSINAMAHLHPASGOLVDPFH
    GRHDLITNLITLISOSRPIIDPTLLRGVRRVRRNYRFOKTANLALATOPALTNAL
    ANVSERIVHELEKLCHEITDPVSSSKIEDILHVMQALIGLTFSSSAHRLSLOEON
    GEPLHMFQAIATVGELEDNMKASLVPPVITAMEORFLIONIEDIOKRLINMIFESDYL
    HKOLYVPEEPLRFYALSSGEMQVLDLYLHQRKQDPLLTGHDMLGKPSPLFK
    ECLLHECEOLKGTLENKODALQFAREFFNHKOPL"
    complement(3074..4309)
    /gene="BH2882"
    complement(3074..4309)
    /gene="BH2882"
    /codon_start=1
    /transl_table=11
    /product="3-oxoacyl-(acyl-carrier protein) synthase"
    /protein_id="BAB06601.1"
    /db_xref="GI:10175504"
    /translation="MNRHYVVTGIGAVPLPGMTAEETWEKATAGOSGVGLITRVEAD
    FPMKVAAYKEDEPTTHIDRKARKMDRTOPAVASALMALDADLEIEDIAPRGV
    WIGSGIGMETIENDRMTIEKGRVSPFPYPMATMASGOVSTITGARINSQSV
    TACASGANSIGAFVYIQGDADVMITGAEAPITIMAVAGFSMAAITSDDPNTAS
    RPEDANRDVGGEGSGILILELSAEKRGAKIYAEIVGATGDAIVYTOPAPEGE
    GAARAKKQALIEDAGLEPDIOYMANHGTSTYNDKXETLAAQVDFDHVNNLAISSTK
    SMTGHLGAAGVEALFEYSKALIEEGIIPTINVEPDPDCLDYVNEARKOGVAVNL
    SNSLGRGHNATVKKYQ"
    complement(4347..5285)
    /gene="BH2883"
    complement(4347..5285)
    /gene="BH2883"
    /codon_start=1
    /transl_table=11
    /product="3-oxoacyl-(acyl-carrier protein) synthase"
    /protein_id="BAB06602.1"
    /db_xref="GI:10175505"
    /translation="MTRAGILIGSYVGEHVYTNKDFEKRLDSDEKIRTRTGIEER
    FAPNDVDSMAEYASALKAGVSAEEDILLVATVTPDMPPTVSTIVOORLCK
    KAAAMDISACAGETIYGLATGOQFIENGQYKHYLVIGVERLSKITDMDNRTLVG
    GAGAAVKGAVSEDSGLIAYELGADGECAMHINQGEYIQMNGREVFKFAVRKMGESAL

```

```

SVLEKAGLSKEDVDLFIHQANIRIMEASRERLEIPVEKMSITTVKRYGNTSSASIPMA
MYDELKDKGIRKDDLLVYFGAGLWMSLALRMGR"
gene
CDS
    5508..5807
    /gene="BH2884"
    5508..5807
    /gene="BH2884"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB06603.1"
    /db_xref="GI:10175506"
    /translation="MSNOERDFFTLMLGRPEESVDVAEENPQESTGVEETKE
    GETQOETMAAPLPLPTFQMEHIFRLAOSLGLKSLSPYKVTIOOLLISQKKD"
    complement(5903..6157)
    /gene="BH2885"
    complement(5903..6157)
    /gene="BH2885"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB06604.1"
    /db_xref="GI:10175507"
    /translation="MNYTGPYRYOYIGTGERFPAVLPLAGLAVGLLEGKRP
    YGTYGPAYGSPYGNPLPTFQMEHIFRLAOSLGLKSLSPYKVTIOOLLISQKKD"
    complement(6226..6555)
    /gene="BH2886"
    complement(6226..6555)
    /gene="BH2886"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB06605.1"
    /db_xref="GI:10175508"
    /translation="MKNKDALRSFYEVKKKKQOLKDKRFEQIOTLPYIELFRQTD
    VAEQVFEVSYIAQLRTGEFEGLRTVEDIKGMELKTEAEVKAEBKLEGLFHLVDLAQL
    AASDKQL"
    complement(6636..7421)
    /gene="BH2887"
    complement(6636..7421)
    /gene="BH2887"
    /note="unknown conserved protein in others"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB06606.1"
    /db_xref="GI:10175509"
    /translation="MGICSSNRQTDMMNAKIYDEHRRFVSAYGEDLIOMLAPKEECV
    LDGCGTDLTEQIHQIGSRVIGVSVSMIRQAKGKPPHDPVOYAEATDLSFSEKFD
    ALEFNAVLHMIKDAEALVITYRSLKPGSRPYAFRGSGNVEITYNTLADTMSQDSR
    WSKTELPMWFFPSIGYTSLSMEVGFRTVHALFPNNPTLLEDGDIRHMLAMAEVYLE
    GKSETERAEIVFEVEAEVDLFDGTINMWADYKRIRIVGMKOL"
    complement(7607..7843)
    /gene="BH2888"

```

Query Match 7.28; Score 84.8; DB 2; Length 294250;  
Best Local Similarity 50.08; Pred. No. 4.4e-09;  
Matches 252; Conservative 0; Mismatches 237; Indels 15; Gaps 1;

```

238 aatcgacgcgcgctgtcgaatcgaacccgctctctcgcgtcgcgaacacgcgct 297
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145487 AGTCGACACGACATTTGGCTTGTATTTAGATGGTGATGGTTAGAAAATCCGAT 145428

298 caattacacgcgtgtgaagtlctcaacgacgcatcgaagcgaacgaacgctagctc 357
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145427 TATTTATGACCTTGTGATGAGTGAAGGCGCTGTGCGCTAAAGCAGCGCTCATCAA 145368

358 tgataaacgcgtgcaaatcgaatcgtctgtcattcaaacctagylcaaatatgccc 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145367 AGATTAACGCTTGAATTTGCAATTTGTTACCTGCTCGGAAAGAAAGATAAGTCTGC 145308

418 tttaggatttgggggaaagccttagtctgctgacatcaaaagatagcgaagaagg 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145307 AGTAACCCCTTAAACAAA-----TGAAGGAAATCTGTCGAGGAGCTTAAACA 145263

```

Qy 478 ttgcccgaatttatactatagatgtctcgaagatgatgacgcggttcaaacgc 537  
 Db 145262 AGAATATGACATGCTTATGATGATGCTGCGGTGGGATCGAACAAGGCTTTAAAAATGC 145203  
 Qy 538 catcacacgcgtaacgaagcgcgtatagttacaacacctgatactacgtcattgagaga 597  
 Db 145202 GGTCCGGGGGGGATGAAAGCATGCTGCTACAACTCTGAATTTCTTCAGTCCGTA 145143  
 Qy 598 tgcagatagatgacagcgtctggaatgataagatgagatataaattgattgt 657  
 Db 145142 TGTGATGCAATTCATTTGGCTTCTCGACAAGAGAGTTGAAGCAACCGCTCTCGTCGT 145083  
 Qy 658 gaacagagttagaactgtttgataaagggtgaagatagtatgtctgtatgttca 717  
 Db 145082 CAATCGAATCCGTGCGCCATATGATGCAAAACGAGAGATGCTGATGTGAGACGATTTGT 145023  
 Qy 718 agagaatttgagattgtatgtatgt 741  
 Db 145022 TTCCATCTTACCATGCACTATT 144999

RESULT 15  
 AEO01931 16486 bp DNA BCT 22-NOV-1999  
 LOCUS Delnoccocus radiodurans R1 section 68 of 229 of the complete  
 DEFINITION  
 AEO01931 AEO00513  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Delnoccocus radiodurans.  
 Bacteria: Thermus/Delnoccocus group; Delnoccocales; Delnoccocus.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 16486)  
 White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
 Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
 Morfitt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
 Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,  
 Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.  
 Genome sequence of the radioresistant bacterium Delnoccocus  
 radiodurans R1  
 Science 286 (5444), 1571-1577 (1999)  
 2 (bases 1 to 16486)  
 White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
 Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
 Morfitt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
 Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,  
 Makarova, K.S., Aravind, L., Daly, M.J., Minion, K.W.,  
 Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,  
 Smith, H.O., Venter, J.C. and Fraser, C.M.  
 Direct Submission  
 Submitted (08-NOV-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA

TITLE  
 JOURNAL  
 FEATURES  
 source  
 Location/Qualifiers  
 1. 16486  
 /organism="Delnoccocus radiodurans"  
 /strain="R1"  
 /db\_xref="taxon:1299"  
 /chromosome="1"  
 /complement(140..427)  
 /gene="DR0751"  
 /note="similar to PID:1001780 percent identity: 53.57;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAI0330.1"  
 /db\_xref="GI:6458462"  
 /translation="MGWFGGRSKETLKDRLVLAVYRAKIAKGVKALRNDLLEY  
 VQKTFPSGNSVVEYEDGKVKVLMASISIEBASISTEEBSAAEDKSDKPSA"  
 complement(427..1257)  
 gene

CDS

/gene="DR0752"  
 /complement(427..1257)  
 /note="DR0752"  
 /note="similar to PID:1001780 PID:1001783 percent  
 identity: 76.89; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="sepium site-determining protein"  
 /protein\_id="AAI0331.1"  
 /db\_xref="GI:6458463"  
 /translation="MPCYAPSPDAKIVYVTSRGVGGKTTTANIGAAALRIGEKVY  
 VIDYDGLRNDLVVWGLSEKRVYFDLVLEKRCRNNOALIRDKRVENLILPASCTRO  
 KDADPEVFEKSVVGLLEEGEDRVLIDSPAGIESGFETAPAAAGALVYVNEVSSV  
 IIVSTNWGEPAVLICKTKAGDAFMATQRIQODVPPFKLTDEEKIWAALIRLFGCS  
 1275..2063  
 /gene="DR0753"  
 1275..2063  
 /gene="DR0753"  
 /note="similar to GB:AL009126 percent identity: 50.23;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAI0332.1"  
 /db\_xref="GI:6458464"  
 /translation="MGVEAGCGVAAACVPGGDSGPECRRREAGRLPGSMTRLRGNONR  
 SVLIIGTLVAALALALALPAVRKSTTTGGEOFTFYANLPYAGGAAVAVNLVGNONR  
 KCPNCKSPFEYVAPLRLKTVGIGKRWMSLYVPLADRLPEDDSKYAQAARCYVYAO  
 GKDAENYTKELIFRAGPETEVMATKSLKSLASLIDIDOKAFPTCLDNDETAQVE  
 2074..2535  
 /gene="DR0754"  
 2074..2535  
 /note="DR0754"  
 /note="similar to GB:AL009126 percent identity: 65.89;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAI0333.1"  
 /db\_xref="GI:6458465"  
 /translation="MNRDRLTYLWALVLAATLGSIFYSEIRHNPCLCAORIFWY  
 PLAVYLGIAAEVGGVGRVRYPLAALGCGATIFONLEWGEVOSIACIVNAAACN  
 2683..3183  
 /gene="DR0755"  
 2683..3183  
 /gene="DR0755"  
 /note="similar to SP:P30529 percent identity: 75.44;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribosomal protein L19"  
 /protein\_id="AAI0334.1"  
 /db\_xref="GI:6458466"  
 /translation="MOTHIKINRGELRGITEDDHNRQLPDRPRGDTVRVDTVREGRN  
 TRSOAPEGVYVAINIGSGSRKSTFVRKISGEVEYVFPAPSPVNDVTVVERGKVRRA  
 KLYVLELNGKARIRKSDRSRYMKDAARQODKANVASQMAAQAQDVIVISAEVVA  
 PETGGG"  
 complement(3254..3826)  
 /gene="DR0756"  
 complement(3254..3826)  
 /gene="DR0756"  
 /note="similar to GP:3449262 percent identity: 52.70;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAI0335.1"  
 /db\_xref="GI:6458467"  
 gene

CDS

gene

gene

CDS

gene

CDS

gene

CDS



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2001, 09:29:10 ; Search time 118.73 Seconds

(without alignments)  
3739,858 Million cell updates/sec

Title: US-09-553-431-3  
Perfect score: 1182

Sequence: 1 aagcttgatcgcacatcc.....tcaaaaaaaaaaaaaaa 1182

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

Database :

N-Geneseq\_36:\*

- 1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	42	3.6	1664976	19 V21209	
2	40.8	3.5	933	21 A08365	Methanococcus jann
3	40	3.4	1049	19 V43610	A. thaliana trans
4	40	3.4	2885	20 200361	Human secreted pro
5	39.6	3.4	1223	20 X25150	Nucleotide sequenc
6	39.6	3.4	6761	20 X20517	Soybean isoflavone
7	39.4	3.3	1217	21 X20146	Polynucleotide seq
8	39.2	3.3	2286	20 X24385	Capisicum annuum (r
C 9	38.9	3.3	2328	20 X99549	Arabidopsis thaliana
10	38.8	3.3	1264	19 V12767	Nucleic acid sequen
11	38.6	3.3	5852	12 Q11710	Metal-regulated tr
C 12	38.6	3.3	9789	17 T41852	Dicystostelium plas
					CDNA encoding plas

13	38.2	3.2	1713	18 T85993	Maize 5-enolpyruv
14	38.2	3.2	1713	18 T86000	Maize 5-enolpyruv
15	38.2	3.2	1713	19 V19277	Maize 5-enolpyruv
16	37.6	3.2	2064	20 X37255	Human gene encodin
C 17	37.6	3.2	1476	20 X13679	Enterococcus faeca
18	37.2	3.1	2781	13 Q13239	HSE CDNA sequence.
19	37.2	3.1	2781	13 Q25712	Sequence of Drosop
20	37.2	3.1	2115	10 N91043	Decay accelerating
21	37.2	3.1	2115	16 Q79863	Decay accelerating
C 22	36.8	3.1	12381	14 Q54144	Sequence of plasmi
23	36.8	3.1	12381	21 Z58381	Streptomyces averm
24	36.4	3.1	574	21 Z51738	Catalpa speciosa t
25	36.4	3.1	1897	17 T32576	Low density lipopr
26	36.4	3.1	1897	20 X88528	Bovine LDL recepto
27	36.4	3.1	1906	17 T32577	Low density lipopr
28	36.2	3.1	1205	18 T92824	Flea saliva protei
29	36.2	3.1	1205	19 V73413	Flea saliva protei
C 30	36.2	3.1	1205	19 V73413	Flea saliva protei
31	36.2	3.1	1559	21 A26377	Human secreted pro
32	36.2	3.0	904	20 X61467	DNA encoding a hum
C 33	36	3.0	1534	20 Z20856	Polynucleotide seq
34	36	3.0	1534	20 Z20856	Sequence encoding
C 35	36	3.0	4590	7 N60472	Sequence of the L.
36	35.8	3.0	1064	21 Z36918	DNA encoding Xenop
C 37	35.8	3.0	1348	21 Z65288	Human secreted pro
C 38	35.8	3.0	3705	20 X77691	Human transmembran
C 39	35.8	3.0	5852	12 Q11710	Dictyostelium plas
C 40	35.8	3.0	8310	20 Z29911	CDNA encoding a SC
41	35.8	3.0	11459	20 X06876	Caenorhabditis ele
C 42	35.6	3.0	857	13 Q23373	CDNA coding for de
43	35.6	3.0	1172	15 Q58669	Protein allergen o
44	35.6	3.0	1502	20 V84577	Human secreted pro
45	35.4	3.0	251	19 X11360	Human diallelic po

#### ALIGNMENTS

RESULT 1  
ID V21209/c V21209 standard; DNA: 1664976 BP.  
XX  
AC V21209;  
XX  
DT 10-NOV-1990 (first entry)  
XX  
DE Methanococcus jannaschli circular chromosome.  
XX  
KW Methanococcus jannaschli; methanogenic archaeon; circular chromosome;  
XX genome; autotrophic; extrachromosomal element; Identification: ds.  
OS  
XX Methanococcus jannaschli.  
PN W09807830-A2.  
XX  
PD 26-FEB-1998.  
XX  
PE 22-AUG-1997; 97MO-US14900.  
XX  
PR 22-AUG-1996; 96US-0024428.  
XX  
PA (GENO-) INSR GENOMIC RES.  
PA (UNIT) UNIT ILLINOIS FOUND.  
PA (UNIT) UNIT JOHNS HOPKINS SCHOOL MEDICINE.  
PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
XX WPI: 1998-169145/15.  
XX Complete genome sequence of methano-genic archaeon, Methanococcus  
XX jannaschli - useful in identification of M. jannaschli genome  
XX fragment

PS Claim 13: Page 152-585; 614pp: English.

XX The present sequence represents the complete 1.66-megabase pair genome  
CC sequence of the Methanococcus jannaschii circular chromosome. The  
CC present invention describes M. jannaschii open reading frames from the  
CC genome sequence. The invention also describes a computer based system  
CC for identifying fragments of the M. jannaschii genome that are  
CC homologous to target nucleotide sequences, comprising: (a) data storage  
CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at  
CC least 99.9% identical to it; (b) search means for comparing a target  
CC sequence to the nucleotide sequence, and (c) retrieval means for obtaining  
CC identity a homologous sequence. The method, which is based on whole genome  
CC the homologous sequence. The method, which is based on whole genome  
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome  
CC of which consists of 3 physically distinct elements, a large circular  
CC chromosome (the 1664976 bp sequence given in V21209), a large circular  
CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a  
CC small circular extra-chromosomal element (the 16550 bp sequence given in  
CC V21211), can be used in the identification of M. jannaschii genome  
CC fragment.

SO Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Query Match 3.6%; Score 42; DB 19; Length 1664976;

Best Local Similarity 47.7%; Pred. No. 2.5; Mismatches 135; Indels 0; Gaps 0;

Matches 123; Conservative 0;

OY 421 agagattg9999aaagcttagttgctgcatgacatgaagaatgagcaagagttg 480

Db 1300845 AGGTTTAAGAAATCAGATTTATTTACTTCAGATGCTTATATGAGTACCTATGA 1300786

OY 481 ccgagatttactactatagatgtccgcagatgatgacgggttcataaccgcat 540

Db 1300785 TTTTGATTATGATTAATTAATGCTCCAGCTGGTTAAATAGAAATGGCTACTCATTT 1300726

OY 541 tacacggcctaagcagccgattatgatacaaccatgatatctactgagagatgc 600

Db 1300725 AGCTATTGCTGATTAACCTTTACTTGTGTCACCCAGAGVGTTCATATTTATTGACGC 1300666

OY 601 agatgagttacagctgctgctggaatgagatgagatgagatgagatgagatg 660

Db 1300665 TGTTAGATTAAAGAAAGTGTGTAATGCTGGAACACCTTAATGGTGTGTGTTAAA 1300606

OY 661 cagagttagaactgatt 678

Db 1300605 TAGGGTTGTAGAGATT 1300588

RESULT 2

A08365 A08365 standard; DNA; 933 BP.

XX A08365;

DT 30-JUN-2000 (first entry)

DE A. thaliana transcription factor DREB1A nucleotide sequence SEQ ID NO:1.

XX Arabidopsis thaliana; transcription factor; DREB1A; DREB2A; DREB1B;

KW DREB1C; DREB2B; plant; stress resistant; dehydration; low temperature;

XX ds.

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 119..769

FT /tag= a

FT /product= "DREB1A"

FT /note= "transcription factor"

XX JP2000060558-A.

XX 29-FEB-2000.

XX 12-AUG-1998; 98JP-0228457.

XX 12-AUG-1998; 98JP-0228457.

XX (NORQ) MORINSUISANSO KOKUSAI NORIN SUIISANGYO.

XX (SEIB-) SEIBUTSOKEI TOKUTEI SANGYO GIJUTSU.

XX WPI: 2000-306709/27.

XX P-PSDB: Y82475.

XX A gene encoding for a transcription factor of plants resistant to

XX stress of dehydration, low temperature or a salt -

XX Claim 4: Page 15-16; 29pp; Japanese.

XX The present invention describes: (1) a protein which binds with a stress

XX responsive element and controls the transcription of a gene in the

XX downstream of the element; (2) a gene encoding for the protein, a

XX recombinant vector having the gene; (3) a transformant having the

XX recombinant vector; (4) a transgenic plant having the gene; (5) a method

XX for determination of the level of stress. The present sequence encodes a

XX transcription factor isolated from Arabidopsis thaliana, designated

XX DREB1A. The transcription factors from the present invention can be used

XX to stress of dehydration, low temperature or a salt.

XX Sequence 933 BP; 273 A; 177 C; 222 G; 261 T; 0 other;

SO Query Match 3.5%; Score 40.8; DB 21; Length 933;

Best Local Similarity 56.8%; Pred. No. 0.13; Mismatches 57; Indels 0; Gaps 0;

Matches 75; Conservative 0;

OY 1051 gataagatttttaattcagtttccagagagaataagatgtagcaactttcat 1110

Db 800 gatacttttatttatttatttatttatttatttatttatttatttatttatttattt 859

OY 1111 ttgataatcattgatttcttgggtcagtgatgatttttactcaaatcaaaaaa 1170

Db 860 ttgtaaaactggaacagcggtgtaaatlaattgattcagttcagttataaaaaaa 919

OY 1171 aaaaaaa 1182

Db 920 aaaaaaaaaa 931

RESULT 3

V43610 V43610 standard; DNA; 1049 BP.

XX V43610;

DT 24-SEP-1998 (first entry)

DE Human secreted protein 10 encoding DNA.

XX Secreted protein; human; cell proliferation; cytokine activity;

KW tissue growth; cellular differentiation; regeneration; activin;

KW inhibitor; chemotactic; haemostatic; thrombolytic; tumour inhibition;

XX anti-inflammatory activity; biomarker; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 21..821

FT /tag= a

FT /product= "human secreted protein"

XX W09825959-A2.

```

XX PD      18-JUN-1998.
XX XX
XX PF      11-DEC-1997;   97MO-US22787.
XX XX
PR      11-DEC-1996;    96US-0032757.
XX XX
PA      (CHIR ) CHIRON CORP.
XX XX
PI      Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
DR      WPt: 1998-348453/30.
DR      p-PDB: W63690.
XX XX

Secreted human polypeptides - having cytokine, cell proliferation or
PT differentiation, activin or inhibin, tumour inhibition or
XX anti-inflammatory activities
XX XX
XX Claim 6; Page 37; 78pp; English.
XX XX

This DNA encodes a human secreted protein. The specification provides
CC secreted protein sequences (W63681 to W63699) invented by the nucleic
CC acid sequences shown in V43601 to V43619. The invention provides a
CC method of identifying a secreted polypeptide which is modified by rough
CC microsome. The secreted proteins can be used in assays to determine
CC biological activities, such as cytokine, cell proliferation, or cellular
CC differentiation activities, tissue growth or regeneration, or cellular
CC inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC thrombolytic activity, receptor/ligand activity, tumour inhibition or
CC anti-inflammatory activity. The proteins can also be used as biomarkers,
CC or to identify tissues or cell types which express the proteins, or a stage-
CC or disease-specific alteration in protein expression. They can be used
CC in protein interaction assays, to identify ligands or binding proteins,
CC or their ability to interact with specific ligands of the secreted proteins
CC using the proteins in screening assays. The proteins and antibodies that
CC bind specifically to the protein can also be used to design diagnostic
CC tests and therapeutic compositions for diseases which may be associated
CC with altered expression of these proteins. Fusion proteins comprising
CC e.g. signal sequences or transmembrane domains of the proteins can be
CC be secreted extracellularly.
CC CC
SO Sequence 1049 BP; 345 A; 214 C; 206 G; 284 T; 0 other:

Query Match          3.4%; Score 40; DB 19; Length 1049;
Best Local Similarity 55.9%; Pctd No. 0.33;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1047 acgtgtaagatggttcaattcagtttcacagagaanaattagatgaaggacctt 1106
DB | ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 891 agtfgtatctttttaagctctcttcacgtgaataagacaacaggsgcttaactactt 950
QY |atttgatcaaatcaattgatctctcttggtagtgatgaatttaccacaatatca 1166
DB 951 tcatctcatcaatcaatataaaacccaattacccttaaataaaaaaaaataaaaaa 1010
QY 1167 aaaaaaaaaaaaaaaaaa 1182
DB | ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1011 aaaaaaaaaaaaaaaaaa 1026

RESULT 4
Z00361 ID Z00361 standard; DNA; 2885 BP.
XX AC Z00361;
DT 26-OCT-1999 (first entry)
XX Nucleotide sequence of human hsfAMP6.

```

[illegible]

Db	1055	acctaattaaaggagatgatttggcttaataattcgatggggaacccgattttcgg	1114
OY	1101	aacttttcattgatcatcaattcgtatcttcttggtcagtgatgaattttactaca	1160
Db	1115	atcttgaatgtgtaacgaggttaacttaattcaattgaagctcgtgttttcaaaa	1174
OY	1161	aatcaaaaaaaaaaaaaaa	1182
Db	1175	aaaaaaaaaaaaaaaaaaaa	1196
RESULT 6			
X20517	ID	X20517 standard; DNA; 6761 BP.	
XX		X20517;	
AC		05-MAY-1999 (first entry)	
XX			
DT			
XX			
DE		Polynucleotide sequence from the genome of Treponema pallidum.	
KW		Treponema pallidum infection; syphilis; Borrelia infection; animal;	
KW		enzyme production; ds.	
XX		Treponema pallidum.	
OS			
XX		WO9859034-A2.	
PN			
XX		30-DEC-1998.	
PD			
XX			
PF		23-JUN-1998; 98WO-US13041.	
XX			
PR		24-JUN-1997; 97US-0050667.	
XX			
PA		(HUMA-) HUMAN GENOME SCI INC.	
PI		Fraser CM;	
XX			
DR		WPI; 1999-081273/07.	
XX			
PT		New isolated Treponema pallidum nucleic acids - used to develop	
PT		products for the detection, diagnosis, characterisation, prevention	
PT		and therapy of T. pallidum infections, particularly syphilis	
XX			
PS		Claim 1; Page 285-289; 1150pp; English.	
XX			
CC		X20500-21243 represent polynucleotide sequences from the genome of	
CC		Treponema pallidum. The sequences can be used for detection,	
CC		diagnosis, characterisation, prevention and therapy for T. pallidum	
CC		infections, particularly syphilis. They can also be used for detecting	
CC		diseases related to Borrelia infections in animals, and for the	
CC		production of biosynthetic products such as enzymes.	
XX			
SQ		Sequence 6761 BP; 1693 A; 1585 C; 1749 G; 1723 T; 11 other:	
Query Match 3.4%; Score 39.6; DB 20; Length 6761;			
Best Local Similarity 53.5%; Pred. No. 0.97; Mismatches 89; Indels 3; Gaps 1			
Matches 106; Conservative 0;			
OY	483	cggaattatactatagatgtgctgcagagatgatgacgggttcataccgcacta	542
Db	2105	cggatcatcatatcatatcgtacaacgagcgtggttcgaagaatgctgaagcttgtg	2164
OY	543	caaccgctaacgaagccggtattagttaacaacccgtatatactgcatgagagcgag	602
Db	2165	catcgtcgcagatgcatgtatgtgtgcacacatgcgcgaacctacgcgaatccacgatgcgt	2224
OY	603	atagatctacagcttgcctgaatgtga--tggaattggagatatataatgattgtga	659
Db	2225	atggaatgataaagaatcatcactgcaactgaggtgataatcgcgatatgaacttgaaagtga	2284
OY	660	acagaagttagaactgatt	677

Db 2285 tagtaaatagatgaatt 2302

RESULT 7  
ID 260146  
XX 260146 standard; cDNA: 1217 BP.  
AC 260146:  
XX  
XX 25-APR-2000 (first entry)  
XX  
XX

DE Capsicum annuum (red pepper) ICC13 gene.

XX Incompatible Capsicum annuum-Colletotrichum gloeosporioides interaction;  
KW ICC13; red pepper; anthracnose disease; fruit crop; antifungal;  
KW antiviral; antibacterial; antinematodal; antiprotzoal; transgenic plant;  
XX pathogen resistance; phytopathogen; ss.

OS Capsicum annuum.

PN US6018038-A.

PD 25-JAN-2000.

XX 04-JAN-1999; 99US-0225244.

XX 04-JAN-1999; 99US-0225244.

PA (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.

PI Oh B, Kostenyuk I, Ko MK;

DR WPI: 2000-136708/12.

DR P-PSDB: Y67379.

XX Isolated cDNA clone from Capsicum annuum, encoding for an incompatible  
PT Capsicum annuum-Colletotrichum gloeosporioides interactions protein,  
PT for producing transgenic plants which exhibit enhanced resistance to  
XX phytophthogens -

PS Claim 1: Fig 1: 15pp; English.

XX This sequence represents a cDNA clone (designated ICC13), which encodes  
CC an incompatible Capsicum annuum-Colletotrichum gloeosporioides  
CC interactions (ICC13) protein. Colletotrichum gloeosporioides is the  
CC causal agent of anthracnose diseases on fruit crops, and shows  
CC incompatible interactions with ripe-red pepper fruit and compatible  
CC interaction with unripe mature green fruit. The ICC13 cDNA is  
CC specifically expressed in the ripe fruit of the pepper, but not the  
CC antimetabolite, ICC13 exhibits antifungal, antibacterial, antiviral,  
CC useful in an expression vector to produce a recombinant DNA expression  
CC system suitable for insertion into cells to form a transgenic plant which  
CC exhibits enhanced resistance to phytophthogens; fungi, bacteria, viruses,  
CC nematodes, mycoplasma like organisms, parasitic higher plants, flagellate  
XX protozoa and insects.

XX Sequence 1217 BP; 391 A; 221 C; 242 G; 363 T; 0 other;

Query Match 3.3%; Score 39.4; DB 21: Length 1217;  
Best Local Similarity 67.9%; Pred. No. 0.51; Mismatches 26; Indels 0; Gaps 0;  
Matches 55; Conservative 0;

QY 1102 actttcatttgatcaatgatttcttggttcagatgaatatttactcaaa 1161  
Db 1125 acttaaatatcatattctctcttgagatcaagatttagatgatttgcgtgaa 1184  
QY 1162 atcaaaaaaaaaaaaaaa 1182  
Db 1185 aaaaaaaaaaaaaaaaaa 1205

RESULT 8

ID X24385 standard; DNA: 2286 BP.

XX X24385;

XX 07-JUN-1999 (first entry)

XX Arabidopsis farnesyl transferase ERA1 gene promoter.

XX Farnesyl transferase; ERA1 gene; transgenic plant;  
KW stress tolerance; cold tolerance; drought tolerance;  
KW salt tolerance; senescence; promoter; guard cell; ss.  
XX Arabidopsis thaliana.

Key Location/Qualifiers  
FT 2284..2296  
FT CDS  
FT /\*tag= a

/\*note= "ERA1 gene start codon"

XX W09906580-A2.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-US15664.

XX 01-AUG-1997; 97US-0054474.

PA (PERF-) PERFORMANCE PLANTS INC.

PI Bonetta D, Cutler S, Ghasseman M, McCourt P,

DR WPI: 1999-153807/13.

XX New isolated Arabidopsis farnesyl transferase gene - used to develop  
PT plants having improved stress tolerance, altered lateral branching,  
PT delayed senescence or altered flowering numbers  
XX Disclosure; Fig 3: 66pp; English.

XX This is the nucleotide sequence of the promoter region of the  
CC farnesyl transferase (FT) ERA1 gene of Arabidopsis thaliana; the  
CC coding region of the gene is provided in X24384, and codes for a  
CC 404-amino acid protein (see W97867) that is involved in the  
CC regulation of lateral branching, regulation of the response to  
CC abscisic acid (ABA) and regulation of senescence. The ERA1  
CC promoter is regulated in the guard cells of the plant and can  
CC affect water loss through the stomates. It can be used for the  
CC expression of gene products in guard cells. A claimed nucleic acid  
CC construct comprises a promoter and a nucleic acid encoding an  
CC inhibitor of a plant FT. Such constructs can be used to produce  
CC plants having improved tolerance to drought, salt and cold stress,  
CC reduced lateral branching, delayed senescence and increased  
CC numbers of flowers. Also claimed in a seed, plant part, cell or  
CC tissue culture or regenerated transgenic plant containing the  
CC claimed nucleic acid construct, the plant being a monocot or dicot,  
CC especially a Brassica sp. A plant having a mutation in the ERA1  
CC gene that results in loss of FT activity is also claimed.

XX Sequence 2286 BP; 791 A; 374 C; 434 G; 681 T; 6 other;

Query Match 3.3%; Score 39.2; DB 20: Length 2286;  
Best Local Similarity 49.7%; Pred. No. 0.77; Mismatches 99; Indels 0; Gaps 0;  
Matches 98; Conservative 0;

QY 986 tgcgtcttggttgatcagagcgtcgtatgtgcaactgaatgtagtgcagc 1045  
Db 1649 ttattattctattctattcttaataataataattgtagatgtagtgcagc 1708  
QY 1046 aagcgtgataagattgtagtcttcagagagaataatggaattgtagcaact 1105

Db 1709 ctattgataaggatgcttaactatcttaataataaacaatgactgttgcctg 1768  
 Oy 1106 ttcatgtgataatcaattgatttcttggttcagtgatgaatttctcaaatca 1165  
 Db 1769 ggcgtacaacgnaattactctcttttttttttgcgaaggagaataataagaagca 1828  
 Oy 1166 aaaaaaaaaaaaaa 1182  
 Db 1829 acagataaacaatcaa 1845

## RESULT 9

X99549/c  
 ID X99549 standard; DNA; 2328 BP.

XX AC X99549;

XX DT 05-OCT-1999 (first entry)

XX DE Nucleic acid sequence from U. urealyticum.

XX KW Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;  
 human urogenital tract; pregnancy; neonatal disease; drug therapy;

XX KM suppurative arthritis; ss.

XX OS Ureaplasma urealyticum.

XX PN W09939007-A1.

XX PD 05-AUG-1999.

XX PF 29-JAN-1999; 99WO-US01972.

XX PR 30-JAN-1998; 98US-0073189.

XX PA (UABR-) UAB RES FOUND.

XX PI Caswell GH, Chen EX, Glass JI, Glass JS, Heiner CR;  
 PI Lefkowitz E;

XX DR WPI: 1999-469343/39.

XX PT Detection of Ureaplasma urealyticum using novel genes, probes and  
 PT primers

XX PS Claim 1: Page 36; 110pp; English.

XX CC The present invention provides methods for the detection and diagnosis  
 CC of Ureaplasma urealyticum infection. It provides novel genes (X99501-681)  
 CC that can be used as a source of primers and probes for the detection and/  
 CC or quantification of U. urealyticum in a biological sample. The probes  
 CC that can be used in the method of the invention by forming target:probe  
 CC complex is complementary to a region selected from one of the 181  
 CC nucleic acid sequences (X99501-681). U. urealyticum is an opportunistic  
 CC pathogen of the human urogenital tract that is a significant cause of  
 CC adverse pregnancy outcome, neonatal disease, and suppurative arthritis.  
 CC As the infections are commonly asymptomatic, it is important to have  
 CC specific and sensitive methods for detecting their presence in a patient.  
 CC Also, as the pathogen has no current antibiotic directed specifically  
 CC against it, it would be advantageous to isolate and detect gene sequences  
 CC which are unique to it, and utilize these as a basis for diagnosis of  
 CC U. urealyticum infection as well as to develop new and improved drug  
 CC therapies. The present invention provides such novel polynucleotide  
 CC sequences (X99501-681).

XX SQ Sequence 2328 BP; 986 A; 290 C; 315 G; 737 T; 0 other;

## Query Match

Best Local Similarity 3.3%; Score 39; DB 20; Length 2328;  
 Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Oy 1029 aatgtatgtgtacagtgatgataagatgttttaattcaagtttcaagagagaaat 1088  
 Db 903 ATTATATATTTTAAAGTAGCAGTTATGATGCTTTGCTTATTTTCAAGAGAT 844  
 Oy 1089 tagaatgttagcaacttcaattgataatcaattcaatttcaatttcaattt 1139  
 Db 843 TAAATCTTTAGCACTTTAGTTAGTTAGTTAAATCTTTAAATTTGATTT 793

## RESULT 10

V12767  
 ID V12767 standard; cDNA; 1264 BP.

XX AC V12767;

XX DT 30-JUL-1998 (first entry)

XX DE Metal-regulated transporter gene IRT2.

XX KW ss; MRT; transgenic plant; pollution; metal-deficiency.

XX OS Arabidopsis thaliana.

XX Key Location/Qualifiers

XX CDS 3..1040  
 /\*tag=a  
 /product="metal-regulated transporter protein IRT2"

XX FT /note="no start codon given"

XX PN W09745000-A1.

XX PD 04-DEC-1997.

XX PF 27-NOV-1996; 96WO-US19065.

XX PR 11-OCT-1996; 96CA-2187728.

XX PR 29-MAY-1996; 96US-0018578.

XX PA (DART-) DARTMOUTH COLLEGE.

XX PA (MINU) UNIV MINNESOTA.

XX PI Elide DJ, Gueriot ML;

XX DR WPI: 1998-032230/03.

XX DR P-PSDB: W41166.

XX PT Novel A. thaliana metal-regulated transporter proteins and related  
 PT DNA - for generating transgenic plants useful for removing  
 PT pollutants from soil or providing nutrient to patients suffering  
 PT from metal-deficiency disorders

XX PS Claim 6: Fig 23; 144pp; English.

XX CC The metal-regulated transporter (MRT) genes (V12763-V12767) encode  
 CC polypeptides which are capable of transporting metals such as Fe(II),  
 CC Pb and Zn. The peptides can be used to generate transgenic plants in  
 CC which the expression of MRT can be altered. These plants can be used  
 CC to remove pollutants from the soil or to treat a disorder associated  
 CC with metal-deficiency. The polypeptides can also be used to identify  
 CC agents which inhibit or stimulate MRT or MRT expression.

XX SQ Sequence 1264 BP; 323 A; 280 C; 261 G; 400 T; 0 other;

Query Match 3.3%; Score 38.8; DB 19; Length 1264;  
 Best Local Similarity 58.8%; Pred. No. 0.75;  
 Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Oy 1069 tcaagtttcaagagaaatagaatgttagcaactttcaattgataatgataatgata 1128  
 Db 1146 tcaagtttcaagagaaatagaatgttagcaactttcaattgataatgataatgata 1205

Oy 1129 ttcttgggtcagtgatgaatttactcaaatcaaaaaaaaaaaaaa 1182

Db 1206 tatcttatgagatgcatgttacttcttaaaaaaaaaaaaaaaaaaaaaa 1259

## RESULT 11

ID Q11710 standard; DNA; 5852 BP.

AC Q11710;

DT 30-JUL-1991 (first entry)

DE Dictyostellum plasmid Ddp2 containing Rep gene.

KW slime mould; replication; Rep gene; ss.

OS Dictyostellum discoideum.

FI Key Location/Qualifiers

FT CDS 2378..5041

FT /tag= a

FT /product= involved in extrachromosomal replication

PN WO9106644-A.

PD 16-MAY-1991.

PP 02-NOV-1990; 90WO-AU00530.

PR 02-NOV-1989; 89AU-0007187.

XX (UYMA-) MACQUARIE UNIV.

XX Slade MB, Chang ACM, Williams KL;

XX WPI: 1991-164194/22.

DR P-PsDB: R11988.

XX Polypeptide facilitating extra-chromosomal replication - of

XX recombinant plasmid in Dictyostellum species

XX Claim 15; Fig 1; 90PP; English.

XX The sequence of Ddp2 has been found to contain the putative open

XX reading frame indicated in the Features table. The possible ORF is

XX flanked by regions with similarity to promoter and polyadenylation

XX signals of known Dictyostellum genes. The RNA and polypeptide

XX product of the Rep gene have not, however, been detected. It is

XX believed that the product is produced in low amounts to positively

XX regulate initiation of plasmid replication. The polypeptide may also

XX contain regions that act as negative regulators of plasmid copy

XX number. See also Q11711 and Q11712.

XX Sequence 5852 BP; 2298 A; 651 C; 708 G; 2195 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 2024 aaaaaaaaaaaaaaaaaa 2040

## RESULT 12

ID T41852/c

AC T41852;

DT 20-FEB-1997 (first entry)

DE cDNA encoding Plasmidium falciparum erythrocyte membrane protein.

DE Plasmidium falciparum; erythrocyte membrane protein; malaria;

DE detection; identification; treatment; prevention; parasite; ss.

OS Plasmidium falciparum MC type.

FI Key Location/Qualifiers

FT CDS 326..9497

FT /tag= a

FT /product= Erythrocyte membrane protein

FT /tag= b

FT /transl\_except= GTR encodes Tyrosine

FT /tag= c

FT /transl\_except= ATW encodes Leucine

FT /tag= d

FT /transl\_except= AAC encodes Aspartic acid

FT /tag= e

FT /transl\_except= GAA encodes Glutamine

FT /tag= f

FT /transl\_except= CCT encodes Arginine

FT /tag= g

FT /transl\_except= AAT encodes Lysine

FT /tag= h

FT /transl\_except= ATA encodes Tyrosine

FT /tag= i

FT /transl\_except= AAC encodes Lysine

FT /tag= j

FT /transl\_except= TTC encodes Isoleucine

FT /tag= k

FT /transl\_except= ATA encodes Histidine

FT /tag= l

FT /transl\_except= ATT encodes Asparagine

FT /tag= m

FT /transl\_except= GCA encodes Tryptophan

FT /tag= n

FT /transl\_except= GCA encodes Tryptophan

FT /tag= n

FT /tag= n

FT /tag= n

FT /tag= n

FT /tag= n



XX  
FA (RHON ) RHONE-POULENC AGROCHIMIE  
XX

Search completed: January 8, 2001, 09:39:00  
Job time: 21409 sec

Tue Jan 9 08:34:57 2001

us-09-553-431-3.rng

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2001, 09:20:42 ; Search time 81.49 seconds

(without alignments)  
2337.606 Million cell updates/sec

Title: US-09-553-431-3  
Perfect score: 1182  
Sequence: 1 aagcttgatcgcacaccc.....tcaaaaaaaaaaaaaaaaaa 1182

Scoring table:

IDENTITY-NUC  
Gapop 10.0, Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

1: /cgn2\_6/pdata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/pdata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/pdata/2/lna/6\_COMB.seq:\*  
4: /cgn2\_6/pdata/2/lna/6\_COMB.seq:\*  
5: /cgn2\_6/pdata/2/lna/6\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.6	4.2	7218	1	US-08-232-463-14
2	45	3.8	7218	1	US-08-232-463-14
3	39.6	3.4	1223	3	US-09-154-874-4
4	39.4	3.3	1217	2	US-09-225-244-1
5	38.8	3.3	1264	2	US-08-758-621-13
6	38.8	3.3	1264	3	US-09-107-658-13
7	38.6	3.3	5852	1	US-07-867-106-2
8	37.2	3.1	2781	1	US-08-178-477B-42
9	36.4	3.1	1897	2	US-08-809-494A-1
10	36.4	3.1	1906	2	US-08-809-494A-1
11	35.6	3.0	857	3	US-08-460-040-1
12	35.6	3.0	1172	1	US-07-945-288-9
13	35.6	3.0	1172	1	US-08-462-831-9
14	35.6	3.0	1172	1	US-08-461-809-9
15	35.6	3.0	1172	1	US-08-461-441-9
16	35.6	3.0	1172	4	PCT-US93-08518-9
17	35	3.0	2295	2	US-08-842-842-6
18	35	3.0	5852	1	US-07-867-106-2
19	34.8	2.9	1422	1	US-08-319-704-5
20	34.8	2.9	2007	3	US-08-747-221B-36
21	34.8	2.9	2007	3	US-08-747-221B-36
22	34.8	2.9	2628	2	US-08-696-944-1
23	34.6	2.9	5261	1	US-08-045-806-3
24	34.6	2.9	5261	1	US-08-365-051B-3
25	34.4	2.9	2026	2	US-08-993-228-3
26	34.4	2.9	4673	4	US-07-638-431-1
27	34.4	2.9	4673	4	PCT-US92-00018-1
28	34.4	2.9	8835	3	US-08-884-324-10

29	34.4	2.9	2894	3	US-08-884-324-14
30	34.2	2.9	2861	1	US-08-299-953-1
31	34.2	2.9	2861	1	US-08-299-953-1
32	34.2	2.9	2861	4	PCT-US95-11231-1
33	34.2	2.9	3881	1	US-08-299-953-2
34	34.2	2.9	3881	1	US-08-459-415-2
35	34.2	2.9	3881	1	PCT-US95-11231-2
36	34.2	2.9	4972	3	US-09-035-648-1
37	34	2.9	2203	1	US-08-076-091C-13
38	34	2.9	2894	1	US-08-285-641-13
39	33.8	2.9	2894	1	US-08-278-091-1
40	33.8	2.9	2894	1	US-08-483-859-1
41	33.8	2.9	2894	1	US-08-472-173-1
42	33.8	2.9	2894	2	US-08-487-167-1
43	33.8	2.9	2894	2	US-08-487-167-1
44	33.8	2.9	2894	2	US-08-487-167-1
45	33.6	2.9	2894	2	US-08-296-149-1
					US-08-801-499-1

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEITLINGER, F.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)836-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-F1s  
US-08-232-463-14

Query Match

4.2%; Score 49.6; DB 1; Length 7218;



SEQUENCE CHARACTERISTICS:  
LENGTH: 1223 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 55..1223  
NAME/KEY: CDS  
LOCATION: 55..990  
US-09-154-874-4

Query Match  
Best Local Similarity 3.4%; Score 39.6; DB 3; Length 1223;  
Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 1041 acagtaagctgtaagagatgttttaattcagatttcagagagaaattagaattgagc 1100  
DB 1055 ACCCTTAATTTAAGAGATGATTTTGCTTATATCGATGGGAACTGGATTTCGG 1114  
OY 1101 aactttcattgatcaattcatttcttcttggtcagtgatgaatttactcaa 1160  
DB 1115 ATCTTGAATGTGAACGACTTTTAACTTATCAATTTAAGCTCTGTTTAAAAA 1174  
OY 1161 aatcaaaaaaaaaaaaaa 1182  
DB 1175 AAAAAAAAAAAAAAAAAA 1196

RESULT 4  
US-09-225-244-1  
; Sequence 1, Application US/09225244  
; Patent No. 6018038  
; GENERAL INFORMATION:  
; APPLICANT: BOUNG-JUN OH  
; APPLICANT: MOON KYUNG KO  
; APPLICANT: IGOT KOSTENYUK  
; TITLE OF INVENTION: Incompatible Plant and Pathogen Interaction Related Gene  
; FILE REFERENCE: 1942/35  
; CURRENT APPLICATION NUMBER: US/09/225,244  
; EARLIER FILING DATE: 1999-01-04  
; EARLIER APPLICATION NUMBER: No. 6018038 Yet Assigned  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Wordperfect 6.1 Windows  
; SEQ ID NO 1  
; LENGTH: 1217  
; TYPE: DNA  
; ORGANISM: Capsicum annuum  
US-09-225-244-1

Query Match  
Best Local Similarity 3.3%; Score 39.4; DB 3; Length 1217;  
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY 1102 actttcattgtaacatgaattgatttcttggtcagtgatgaatttactcaa 1161  
DB 1125 acttaaatatataatttctcttggaagatcaaaagtttgaagatgagttgtgaaa 1184  
OY 1162 aatcaaaaaaaaaaaaaa 1182  
DB 1185 aaaaaaaaaaaaaaaaaa 1205

RESULT 5  
US-08-758-621-13  
; Sequence 13, Application US/08758621  
; Patent No. 5846821  
; GENERAL INFORMATION:

APPLICANT: Guerino, Mary Lou, and Eide, David J.  
TITLE OF INVENTION: Metal-Regulated Transporters and Uses Therefor  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758,621  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/018,578  
FILING DATE: 29-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silver, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: DCI-099CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1264 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..1037  
US-08-758-621-13

Query Match  
Best Local Similarity 3.3%; Score 38.8; DB 2; Length 1264;  
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 1069 tcaagttcagagagaaattagaattgtagcaactttcatttgatcaattcgtga 1128  
DB 1146 TTAGTTTGTATAGTCATGTGAAGCGTTTGAAGAATTTCTTTATGACATTTGTAATT 1205  
OY 1129 ttctttggtcagtgatgaatttctactcaaaatcaaaaaaaaaaaaaa 1182  
DB 1206 TATTTTATGATGCGATGTTTACTTCTTAAAAAATAAAAAAAAAAAAAA 1259

RESULT 6  
US-09-107-858-13  
; Sequence 13, Application US/09107858  
; Patent No. 6162900  
; GENERAL INFORMATION:  
; APPLICANT: Guerino, Mary Lou et al.  
; TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR  
; FILE REFERENCE: DCI-099CPDV  
; CURRENT APPLICATION NUMBER: US/09/107,858  
; EARLIER FILING DATE: 1998-06-30  
; EARLIER APPLICATION NUMBER: 08/758,621  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 1264  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:

QY 1093 attgtagcaacttttcatttgatcaatccaattgtattctcttggttcagtgatgaattt 1152



```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/162,722
FILING DATE: 8-NOV-93
APPLICATION NUMBER: 07/458,642
FILING DATE: 13-FEB-90
APPLICATION NUMBER: PCT/AU88/00195
FILING DATE: 17-JUNE-88
APPLICATION NUMBER: PI 2523/87
FILING DATE: 18-JUNE-87
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMT-021CN2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 857 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..736
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 70..736

```

```

Query Match Similarity      3.0%; Score 35.6; DB 3; Length 897;
Best Local Similarity 57.0%; Pred. No. 0.63;
Matches      65; Conservative      0; Mismatches      49; Indels      0; Gaps      0;

Qy      1069   tcagtttcaaggagaagaatttgaaatgtgacgaacttcattgtgatcattcaattgtga      1128
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       728   TCATTCCTGTAACAAAAGACAATTCTTATATGATGTGCACATAATTATTTAAATCAAA      787

Qy      1129   ttctcttgttccagtgatgatatttctaccacaatacaaaaaaaaaaaaaaaa      1182
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       788   AATTTTGAAGAAATGAAATAAATTCATTCCACCAAAATTTAAAAAAAAAAAAAAAAA      841

RESULT      12
US-07-945-288-9
; Sequence 9, Application us/07945288
; Patent No. 543948
; GENERAL INFORMATION: Wayne R.

```

APPLICANT: CHUA, KAY  
TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM  
TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, SUITE 510  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,288  
FILING DATE: 19920910  
CLASSIFICATION: 314  
PRIOR APPLICATION DATA:

RESULT 14  
 US-08-461-809-9  
 : Sequence 9, Application US/08461809  
 Patent No. 5770202  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM  
 NUMBER OF INVENTION: DERMATOPHAGOIDES  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 STATE STREET, SUITE 510  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII TEXT  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/461,809  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/945,288  
 FILING DATE: 10 SEPTEMBER 1992  
 APPLICATION NUMBER: US 580,655  
 FILING DATE: 11 SEPTEMBER 1990  
 APPLICATION NUMBER: US 458,642  
 FILING DATE: 13 FEBRUARY 1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MANDRAGOURAS, AMY E.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1172 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single

US-08-461-441-9

```
Search completed: January 8, 2001, 09:20:56
Job time: 20531 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2001, 02:32:40 ; Search time 1361.86 Seconds

(without alignments)  
5032.326 Million cell updates/sec

Title: US-09-553-431-1  
Sequence: 1 atggcgctctctgagatgtt.....tctctctctcttgagcgc 978

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est25:*
26:	gb_est26:*
27:	gb_est27:*
28:	gb_est28:*
29:	gb_est29:*
30:	gb_est30:*
31:	gb_est31:*
32:	gb_est32:*
33:	gb_est33:*
34:	gb_est34:*
35:	gb_est35:*
36:	gb_est36:*
37:	gb_est37:*
38:	gb_est38:*
39:	gb_est39:*
40:	gb_est40:*
41:	em_est41:*
42:	em_estfun:*
43:	em_esthum1:*

44:	em_esthum2:*
45:	em_esthum3:*
46:	em_esthum4:*
47:	em_esthum5:*
48:	em_esthum6:*
49:	em_esthum7:*
50:	em_esthum8:*
51:	em_esthum9:*
52:	em_esthum10:*
53:	em_esthum11:*
54:	em_esthum12:*
55:	em_esthum13:*
56:	em_esthum14:*
57:	em_esthum15:*
58:	em_esthum16:*
59:	em_esthum17:*
60:	em_esthum18:*
61:	em_esthum19:*
62:	em_esthum20:*
63:	em_esthum21:*
64:	em_esthum22:*
65:	em_esthum23:*
66:	em_esthum24:*
67:	em_esthum25:*
68:	em_esthum26:*
69:	em_esthum27:*
70:	em_esthum28:*
71:	em_esthum29:*
72:	em_esthum30:*
73:	em_esthum31:*
74:	em_esthum32:*
75:	em_esthum33:*
76:	em_esthum34:*
77:	em_esthum35:*
78:	em_esthum36:*
79:	em_esthum37:*
80:	em_esthum38:*
81:	em_esthum39:*
82:	em_esthum40:*
83:	em_esthum41:*
84:	em_esthum42:*
85:	em_esthum43:*
86:	em_esthum44:*
87:	em_esthum45:*
88:	em_esthum46:*
89:	em_esthum47:*
90:	em_esthum48:*
91:	em_esthum49:*
92:	em_esthum50:*
93:	em_esthum51:*
94:	em_esthum52:*
95:	em_esthum53:*
96:	em_esthum54:*
97:	em_esthum55:*
98:	em_esthum56:*
99:	em_esthum57:*
100:	em_esthum58:*
101:	em_esthum59:*
102:	em_esthum60:*
103:	em_esthum61:*
104:	em_esthum62:*
105:	em_esthum63:*
106:	em_esthum64:*
107:	em_esthum65:*
108:	em_esthum66:*
109:	em_esthum67:*
110:	em_esthum68:*
111:	em_esthum69:*
112:	em_esthum70:*
113:	em_esthum71:*
114:	em_esthum72:*
115:	em_esthum73:*
116:	em_esthum74:*

117: em\_estp16:\*  
118: em\_estp17:\*  
119: em\_estp18:\*  
120: em\_estp19:\*  
121: em\_estp20:\*  
122: em\_estp21:\*  
123: em\_estp22:\*  
124: em\_estp23:\*  
125: em\_estp24:\*  
126: em\_estp25:\*  
127: em\_estp26:\*  
128: em\_estp27:\*  
129: em\_estp28:\*  
130: em\_estp29:\*  
131: em\_estp30:\*  
132: em\_estp31:\*  
133: em\_estp32:\*  
134: em\_estp33:\*  
135: em\_estp34:\*  
136: em\_estp35:\*  
137: em\_estp36:\*  
138: em\_estp37:\*  
139: em\_estp38:\*  
140: em\_estp39:\*  
141: em\_estp40:\*  
142: em\_estp41:\*  
143: em\_estp42:\*  
144: em\_estp43:\*  
145: em\_estp44:\*  
146: em\_estp45:\*  
147: em\_estp46:\*  
148: em\_estp47:\*  
149: em\_estp48:\*  
150: em\_estp49:\*  
151: em\_estp50:\*  
152: em\_estp51:\*  
153: em\_estp52:\*  
154: em\_estp53:\*  
155: em\_estp54:\*  
156: em\_estp55:\*  
157: em\_estp56:\*  
158: em\_estp57:\*  
159: em\_estp58:\*  
160: em\_estp59:\*  
161: em\_estp60:\*  
162: em\_estp61:\*  
163: em\_estp62:\*  
164: em\_estp63:\*  
165: em\_estp64:\*  
166: em\_estp65:\*  
167: em\_estp66:\*  
168: em\_estp67:\*  
169: em\_estp68:\*  
170: em\_estp69:\*  
171: em\_estp70:\*  
172: em\_estp71:\*  
173: em\_estp72:\*  
174: em\_estp73:\*  
175: em\_estp74:\*  
176: em\_estp75:\*  
177: em\_estp76:\*  
178: em\_estp77:\*  
179: em\_estp78:\*  
180: em\_estp79:\*  
181: em\_estp80:\*  
182: em\_estp81:\*  
183: em\_estp82:\*  
184: em\_estp83:\*  
185: em\_estp84:\*  
186: em\_estp85:\*  
187: em\_estp86:\*  
188: em\_estp87:\*  
189: em\_estp88:\*

190: gb\_gss25:\*  
191: gb\_gss26:\*  
192: gb\_gss27:\*  
193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	408.4	41.8	457	140	C99848	C99848 C99848 YAC
2	408.2	41.7	710	93	AM668240	AM668240 GA_Ea001
3	341.4	34.9	873	94	AW729746	AW729746 GA_Ea002
4	214.4	21.9	557	107	BE359432	BE359432 DGL_40_A0
5	190.6	19.5	456	28	AL370067	AL370067 MRA35D01
6	181	18.5	276	174	B23006	B23006 F10H1F_1GF
7	169.2	17.3	588	107	BE359345	BE359345 DGL_40_A0
8	165	16.9	278	40	BE118872	BE118872 SCL083.D0
9	161.8	16.5	362	104	BE124676	BE124676 pblt12 so
10	120.4	12.3	104	165	AO358203	AO358203 LERAW015R
11	100.8	10.3	386	92	AW574469	AW574469 GSSB194
12	68.6	7.0	463	170	A2303004	A2303004 GSSB194
13	51	5.2	463	170	A2303004	A2303004 GSSB194
14	50	5.1	548	142	N37510	N37510 18737 Lambda
15	49.4	5.1	508	37	AV527004	AV527004 AV527004
16	47.2	4.8	1101	190	CNS010B7	AL098761 Drosophila
17	46.6	4.5	583	108	BE453489	BE453489 894071A05
18	44.4	4.5	880	190	CNS00KMD	AL078160 Drosophila
19	43.6	4.5	1101	190	CNS00001	AL065414 Drosophila
20	43.6	4.4	1101	190	CNS01844	AL108862 Drosophila
21	42.2	4.3	588	105	BE194224	BE194224 HVSME008
22	40.4	4.1	959	190	CNS004XY	AL055406 Drosophila
23	39.6	4.0	939	190	CNS000CMG	AL064643 Drosophila
24	39.6	4.0	1101	190	CNS0182P	AL059400 Drosophila
25	39.6	4.0	507	108	BE462584	AL108811 Drosophila
26	39.4	4.0	551	93	AM626014	BE462584 EST324870
27	39.4	4.0	420	107	BE341746	AM626014 EST319921
28	39.2	4.0	1014	135	BE747742	BE341746 EST394577
29	38.8	4.0	1101	190	CNS0106X	BE747742 601578717
30	38.8	3.9	386	37	AV554133	AL098595 Drosophila
31	38.6	3.9	1101	190	CNS017YX	AV554133 AV554133
32	38.4	3.9	1101	190	AO543931	AL108171 Drosophila
33	37.4	3.8	478	159	AO543931	AO543931 RPT-11-3
34	37.4	3.8	1021	190	CNS00DLS	AL072703 Drosophila
35	37.4	3.8	1221	190	CNS01690	AL106476 Drosophila
36	37.2	3.8	1101	190	CNS00396	AL065921 Drosophila
37	37	3.8	453	96	AW922683	AW922683 DGL_45_D0
38	37	3.8	561	94	AW746558	AW922683 WS1_54_A0
39	37	3.8	578	93	AW680908	AW746558 WS1_54_A0
40	37	3.8	581	97	AW924763	AW680908 WS1_8_A11
41	37	3.8	845	137	BE918164	AW924763 WS1_72_G0
42	37	3.8	845	190	CNS013KO	BE918164 OVI_2_G02
43	37	3.8	1080	190	CNS00EPP	AL102978 Drosophila
44	37	3.8	1101	190	CNS00LOO	AL069494 Drosophila
45	37	3.8	1101	190	CNS0100X	AL068607 Drosophila

## ALIGNMENTS

RESULT 1  
C99848 457 bp mRNA EST 08-OCT-1998  
LOCUS C99848 YAC clone C1C8B11 region-specific CDNA Arabidopsis thaliana  
DEFINITION C99848 C99848  
ACCESSION C99848  
VERSION C99848.1 GI:3719205  
KEYWORDS EST.  
SOURCE tlaie cress.

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 457)  
Ito, T.  
Regional insertional mutagenesis of genes on Arabidopsis thaliana  
chromosome V using Ac/Ds transposon in combination with a cDNA  
scanning method  
Unpublished (1998)

JOURNAL

Contact: Takuya Ito  
Laboratory of Plant Molecular Biology  
The Institute of Physical and Chemical Research (RIKEN)  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Email: tallo@rcc.riken.go.jp.  
Location/Qualifiers  
1. 457

FEATURES

source  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone\_lib="YAC clone CTC811 region-specific cDNA"

BASE COUNT 112 a 125 c 96 g 122 t 2 others

ORIGIN

Query Match 41.8%; Score 408.4; DB 140; Length 457;  
Best Local Similarity 99.3%; Pred. No. 7.2e-117;  
Matches 409; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atggagctctgagatgtgtctcaagaatcatcatctctctccatcatctc 60  
DB 46 atggcgtctgagatgtgtctcaagaatcatcatctctctccatcatctc 105  
QY 61 tcaaaaaagacttaataatctcccaagaatctgcaaaccttagacagagatc 120  
DB 106 tccaaaaagcctctaatatcttcccaagaatctgcaaaccttagacagagatc 165  
QY 121 atagagctcgttcttaattatcgcaaacctgagcagcagcagcagcagc 180  
DB 166 atagagctcgttcttaattatcgcaaacctgagcagcagcagcagcagc 225  
QY 181 gtcgtatcaaccctcggaagagcagcagcagcagcagcagcagcagcagc 240  
DB 226 gtcgtatcaaccctcggaagagcagcagcagcagcagcagcagcagcagc 285  
QY 241 cctctctcgtctgcttaagcttctcaagctgctgcaatgagcagcagcagc 300  
DB 286 cctctctcgtctgcttaagcttctcaagctgctgcaatgagcagcagcagc 345  
QY 301 aacctgcatctcctctctagagcttagaagaatcgatcaatcaacttgctgaggtata 360  
DB 346 aacctgcatctcctctctagagcttagaagaatcgatcaatcaacttgctgaggtata 405  
QY 361 aacggaagatctgctgctgctgctgctgctgctgctgctgctgctgctg 412  
DB 406 aacggaagatctgctgctgctgctgctgctgctgctgctgctgctgctg 457

RESULT 2

AM668240 710 bp mRNA EST 06-APR-2000  
LOCUS GA\_Ea0013E22 Gossypium arboreum 7-10 dpa fiber library Gossypium  
DEFINITION GA\_Ea0013E22 Gossypium arboreum 7-10 dpa fiber library Gossypium  
ACCESSION AM668240  
VERSION AM668240.1 GI:7502620  
KEYWORDS EST.  
SOURCE Gossypium arboreum.  
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;  
Malvales; Malvaceae; Gossypium.  
REFERENCE 1 (bases 1 to 710)  
AUTHORS Leslie, A., Fritsch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.  
TITLE An integrated analysis of the genetics, development, and evolution

JOURNAL  
COMMENT  
of the cotton fiber  
unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu

FEATURES

source  
Location/Qualifiers  
1. 710

/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="0400"  
/db\_xref="taxon:29729"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/issue\_type="Fibers isolated from bolls harvested 7-10  
dpa"

BASE COUNT 185 a 157 c 213 g 155 t

ORIGIN

Query Match 41.7%; Score 408.2; DB 93; Length 710;  
Best Local Similarity 76.6%; Pred. No. 9.6e-117;  
Matches 500; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 118 ccaatcagctcgttcttaattatcgcaaacctgagcagcagcagcagc 177  
DB 57 ccaatcagctcgttcttaattatcgcaaacctgagcagcagcagcagc 116  
QY 178 atcgtctgctatcaaccctcggaagagcagcagcagcagcagcagcagc 237  
DB 117 gtcgtctgctatcaaccctcggaagagcagcagcagcagcagcagc 176  
QY 238 gctctctcgtctgctgctgctgctgctgctgctgctgctgctgctg 297  
DB 177 gctctctcgtctgctgctgctgctgctgctgctgctgctgctgctg 236  
QY 298 cgtaacctcgtatctcctctctagagcttagaagaatcgatcaatcaacttgctgaggt 357  
DB 237 cgtaacctcgtatctcctctctagagcttagaagaatcgatcaatcaacttgctgaggt 296  
QY 358 ataacgagagatctgctgctgctgctgctgctgctgctgctgctgctg 417  
DB 297 cgtaacgagagatctgctgctgctgctgctgctgctgctgctgctgctg 356  
QY 418 gaatgctatgataatcaaaccttagaagaatcgatcaatcaacttgctgaggt 477  
DB 357 gaatgctatgataatcaaaccttagaagaatcgatcaatcaacttgctgaggt 416  
QY 478 ttggaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 537  
DB 417 ttggaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 476  
QY 538 atcgaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 597  
DB 477 atcgaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 536  
QY 598 gcaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 657  
DB 537 gcaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 596  
QY 658 ttgttagaagcgaatcgatcaatcaaaccttagaagaatcgatcaatcaacttgctgaggt 717  
DB 597 ctgttagaagcgaatcgatcaatcaaaccttagaagaatcgatcaatcaacttgctgaggt 656  
QY 718 atgataaagagagatgctgctgctgctgctgctgctgctgctgctgctg 770  
DB 657 atgataaagagagatgctgctgctgctgctgctgctgctgctgctgctg 709

```

RESULT 3
LOCUS   AM729746
DEFINITION  GA_Ea0025013 Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum CDNA clone GA_Ea0025013, mRNA sequence.
ACCESSION  AM729746
VERSION    AM729746.1
KEYWORDS   GI:7627348
SOURCE     Gossypium arboreum.
ORGANISM   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II;
            Malvales; Malvaceae; Gossypium.
REFERENCE  1 (bases 1 to 873)
AUTHORS   Leslie, A., Frisch, D., Yu, Y., Wood, T.C., Wing, R.A. and Wilkins, T.A.
TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL   Unpublished (2000)
COMMENT   Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            High quality sequence stop: 873.
FEATURES
    source
        1..873
        /organism="Gossypium arboreum"
        /strain="AKA"
        /cultivar="8400"
        /db_xref="taxon:29729"
        /clone="GA_Ea0025013"
        /tissue_type="Fibers isolated from bolls harvested 7-10
        dpa"
        /lab_host="E. coli"
        /note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT  216 a 186 c 268 g 203 t
ORIGIN
Query Match          34.9%; Score 341.4; DB 94; Length 873;
Best Local Similarity 72.9%; Pred. No. 8.6e-96;
Matches 466; Conservative 0; Mismatches 171; Indels 2; Gaps 2;

OY 118 ccaatacgaatcggtcttcaatttaacgaacacggaactcgccgagagaacgcgcgt 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 56 cccatttcttgccttgatgaatataaacctgacgtgacgagcgggagacacacgc 115

OY 178 atcgctgtatataccctccggaagacggtgttg-aaagacgaacacacgcgaatgt 236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 ggcgtgttatatacctccggaagacggtgttg-aaagacgaacacacgcgaatgt 175

OY 237 cggctctctcgcgtcgttaacggttcctcagttgttcgcatgacgcgcgaactgtct 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 ggctcttcttgcgtcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 235

OY 297 ccgtaacct-cgaatctctcctcaggttagaagacgagtcgaatctacactgctgcaag 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 acataacacatgacgtctgttgccttgagaaatccgtgaactatgctgctgagag 295

OY 356 ttataacgagatgtgtcgtcgaatcaactcgtgacgtgataagcgttgcaatt 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 296 tactcatgacgagtgccgactgacacatgattggaagatgaagcgttgcaact 355

OY 416 tcgaattcatatatacctcaaacctagaactcgaatgagatttggtgtaag 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 tatataatgcttgcatttccaaagccagatcgaaatcaccatggtgattgaggaag 415

OY 476 catggaatgagcttgatgagcttgaaactagaccggaaggttcaacggaattatca 535

```

```

DB 416 cgttggtttgctagtgacgacctgaatgcccggagaaagggctacacggatttcataac 475
OY 536 tcatcagatgtcctcgcgaagatcgatccggaattcataaccgcaactcctccgcgaat 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 476 taataatgttgacgacggcgatgacgacgacgacgacgacgacgacgacgacgacgac 535
OY 596 aagacatctcgtgtacacaccccgatataacacgcttaagagatgataaggttaccg 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 536 agcgcgttttggtgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 595
OY 656 gtctgtgaagatcgatgaaatcagagataataaagatgtgtgacagagtgagactg 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 ggcgtttgaatgagatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 655
OY 716 atatgattaaagagagagatgattgattgattgattgattgattgattgattgattgatt 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 ctctgatttagagggagagatgattgattgattgattgattgattgattgattgattgatt 694

RESULT 4
LOCUS   BE359432
DEFINITION  Dcl_40_A03.91_A002 Dark Grown 1 (Dg1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION  BE359432
VERSION    BE359432.1
KEYWORDS   GI:9300989
SOURCE     EST.
ORGANISM   sorghum.
            Sorghum bicolor
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
            1 (bases 1 to 557)
REFERENCE  1 (bases 1 to 557)
AUTHORS   Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt
            L.H.
TITLE     An EST database from Sorghum: dark-grown seedlings
JOURNAL   Unpublished (2000)
COMMENT   Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmp@prattuga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolYTmix
            High quality sequence start: 6
            High quality sequence stop: 531
            POLYA-No.
FEATURES
    source
        1..557
        /organism="Sorghum bicolor"
        /db_xref="taxon:4558"
        /clone="Organ: 5-day-old dark-grown seedlings; Vector:
        lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
        made from poly-A RNA in the cloning vector lambda Zap II.
        Clones to be sequenced were prepared by mass excision."
BASE COUNT  116 a 127 c 174 g 140 t
ORIGIN
Query Match          21.9%; Score 214.4; DB 107; Length 557;
Best Local Similarity 69.7%; Pred. No. 4.4e-56;
Matches 290; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

OY 523 ccgattatcatcatcattgtcctcgcaggaatcgatgacggaattcataaccgcaatt 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 cccgccttcattcattcattgacgtgcccccgcagaggttgatgccgggttgcctgacgtcatt 60

OY 583 actcggcgaatgaagcagcttctgtaacaactccgatatataacagcgttaagagatgct 642

```

Db 61 GCACCTGACAGAGAGCGAGTCTGACACCCCTGACATTACGGCTCTCCGATGCT 120  
 QY 643 gataaggttcaaggttctgtagaattgcaatgcaagaatataagaatgatgtagac 702  
 Db 121 GACCGTCTCCAGACGCTTGAGCTGCGATGCGATCAAAATATCAAGATTATGTCAAC 180  
 QY 703 agagtagaactatgatataagaagaagatagatgtagatgtagatgtagagag 762  
 Db 181 CAGAGTCCGCCAGACCTGCTGAAGGGGAGACATGATGATCAGACTTATGTCAGAA 240  
 QY 763 atgttgagattgcatgtgtgtgtaattcctgaagatcgaagttatcgaagaag 822  
 Db 241 ATGCTTGGGGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 QY 823 aatcgaaggttccgctgtgttctgaataagcctcctgaagctcgaagttatcgaagaag 882  
 Db 301 AATAGCGGTGATACATTTGCTGCTCAACACCGCCGCGCCGCTGCTGCTGCTG 360  
 QY 883 cagcggcttggagactcgttgaagaagaatagatgaagcgtgtgaagtagaaga 938  
 Db 361 CAGGCTACTTGGCGATTGCTGTAAGAGATGATGATGATGATGATGATGATGATG 416

RESULT 5  
 LOCUS AL370067 456 bp mRNA EST 03-AUG-2000  
 DEFINITION MIBA35D01F1 MIBA Medicago truncatula cDNA clone MIBA35D01 T3, mRNA  
 sequence.  
 ACCESSION AL370067  
 VERSION AL370067  
 KEYWORDS EST  
 SOURCE GI:9669820  
 ORGANISM barrel medic.  
 Medicago truncatula  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 Fabales; Fabaceae; Papilionoideae; Medicago.  
 1 (bases 1 to 456)  
 JOURNAL  
 AUTHORs  
 REFERENCE  
 COMMENT

JOURNAL  
 COMMENT  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex France  
 Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Contact: Pascal Gamas and Etienne-Pascal Journet, Laboratoire de  
 Biologie Moleculaire des Relations Plantes-Microorganismes,  
 CNRS-INRA, BP 27 33326 Castanet-Tolosan Cedex, France (Email :  
 http://sequence.toulouse.inra.fr/Mtruncatula.html).  
 Location/Qualifiers  
 1. 456  
 /organism="Medicago truncatula"  
 /cultivar="Jemalong"  
 /db\_xref="taxon:3880"  
 /clone="MIBA35D01"  
 /clone\_1b="MIBA"  
 /tissue\_type="root tips"  
 /dev\_stage="harvested after 3 days of N-starvation"  
 /note="Vector: pBluescript PSK; Site\_1: EcoRI; Site\_2:  
 XhoI; Plants were grown in an aeroponic chamber for 14  
 days on nitrogen-rich medium followed by 3 days on N-free  
 medium. RNA was extracted from root tips (1-3 cm). cDNA  
 was prepared from polyA+ enriched RNA. The cDNA was  
 directionally ligated into uni-zapR vector from  
 Strategene and packaged using GigaPack Gold packaging  
 extracts. Plasmids containing cDNA inserts were  
 mass-excised from phage stocks using ExsacII helper phage  
 and propagated in SOLR cells. Clone ordering and  
 sequencing was performed by the Centre National de  
 Sequencage (Genoscope, Evry, France)."

BASE COUNT 110 a 68 c 141 g 137 t  
 ORIGIN

Query Match 19.5% Score 190.6; DB 28; Length 456;  
 Best Local Similarity 81.8%; Pred. No. 1,2e-48;  
 Matches 220; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 517 ggtcaccggaattcatalcatalcatalcgttcctcgaagaatgaltccggaattcataacc 576  
 Db 188 GGTCTGTCGGGACTTATTCCTGATGATGATGATGATGATGATGATGATGATGATG 247  
 QY 577 gccattaccgcgcgaatgaagcagcttctgtgaacaactccgataatacagcgttaag 636  
 Db 248 GCGATTACACCGCTATATGAAGGGTTTGTATTAACAACCCGGATATTAAGTGGTTGAG 307  
 QY 637 gatgcctgaagaaggttgaaggttctgtagaattcgaatggaatcgaagaatgaatgalt 696  
 Db 308 GATGCGGATGAGAGTGCAGCGGATTTGAGCTGTGATGCGATTTAGGATATTAAGATGATA 367  
 QY 697 gtgaacaagatgagaactatgatatgaagaagaagatagatgtagcagtttagatgtg 756  
 Db 368 GTGATAGGTTAGGACGCGATATGATTAAGGTGAAGATATGATGATGATGATGATGATG 427  
 QY 757 cagcgaatgttggagattgcatgtctgtg 785  
 Db 428 CAGAGATGTTGGGTTTCCGCTTGGCTTGG 456

RESULT 6  
 LOCUS B23006/C 276 bp DNA GSS 10-OCT-1997  
 DEFINITION F16H1TF IGF Arabidopsis thaliana genomic clone F16H1, DNA sequence.  
 ACCESSION B23006  
 VERSION B23006  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 276)  
 JOURNAL  
 AUTHORs  
 REFERENCE  
 COMMENT

JOURNAL  
 COMMENT  
 Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Coffey, E., Golden, K.,  
 Johnson, K., Adams, M.D. and Venter, D.C.  
 A BAC End Sequence Database for Identifying Minimal Overlaps in  
 Arabidopsis Genomic Sequencing  
 Unpublished (1997)  
 Other GSSs: F16H1R  
 Contact: Steve Rounsley  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: rounsley@tigr.org  
 Seq primer: M13-21  
 Class: BAC ends  
 High quality sequence stop: 276.  
 Location/Qualifiers  
 1. 276  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="F16H1"  
 /clone\_1b="IGF"  
 /sex="thermophilodite"  
 /note="Vector: BelobAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
 Produced by Thomas Altmann"

BASE COUNT 77 a 76 c 53 g 70 t  
 ORIGIN

Query Match

18.5%; Score 181; DB 174; Length 276;

[illegible]

Query Match	16.98;	Score 165;	DB 108;	Length 773;
Best Local Similarity	60.18;	Pred. No. 1.5e-40;		

Matches	326:	Conservative	0:	Mismatches	211:	Indels	5:	Gaps	3:
QY	151	ccggagcactcgccggagaaagcgcggtatcgtcgtatcaactccggaaagcggtgtc	210						
Db	99	CCGGAGCTGTCCGGCCCGCCGCGCGGGGTGGTGGTGATCCTCGGGCAAGGGCGCGGTG	158						
QY	211	ggaagaagacacacacacgcaaalbgtcgtctctctctcgtcgtcttctacggttctcagtc	270						
Db	159	GGGAGACACACACACACCGCCCAACTGGCCCGCCCTCCGTGGCGCGGTGGGGCTCCCGCGTG	218						
QY	271	gtcgscatttgcgcgcgaacttgcgtctcgttaacctcgtctctctctcttcttgggttgaagaat	330						
Db	219	GTGGCCCGTGGAGCGCGAGCGGGGCTCCGGAACCTCGACCTGCTCTGGGCTGGAGAAC	278						
QY	331	cgaagtcaattcaactcttgcgtcgaagttataacgagaaattgctcgtcgtatcaagctctg	390						
Db	279	CGGGTGAACCTGTACCCGCGCGAGCTGCTCGCGCGAGACTGGCGGCTCGACACAGCCCTTC	338						
QY	391	gtacgttgaatagcgttctgtcaatttctgaatttgcgtatgatalatctaaacctatgacgaa	450						
Db	339	GTCCGCCACACCGCTCCCTCCGGGGGCTCCACCTGCTCTGTGCTTCCAGGCCCGCTCCAG	398						
QY	451	cttcgcgatggaatttgttgttaagcatttgaaatggtcttgtatgctgttgaanaactaga	510						
Db	399	CTGCCCTTCGCGCTTGCGCTCCAAAGCCCTTCACCTGGTGGCGGAGGCCCTTCGCGCGGCG	458						
QY	511	ccggaaggttccacggatttcatctatctatctgtctctcgtcgaagaatgattccggaatc	570						
Db	459	GCCGA---CCCGCCCGCTTCATNCTCATCGACTGCCCTGCAGGCTTGATGCTCGGATT	515						
QY	571	ataaccgcaattactcggcggaatgaagcaattctgttaacaactcggatataaacgacgc	630						
Db	516	GTCACAGCCATTCGCCCTTCGACGAAC-ACCAGGCTGTGGTACCACATCCTGACATCAC-AGA	573						
QY	631	ttaaggaatgctcgtataggttccaggttgtttagaatcgatcgatcgatcgatgaataag	690						
Db	574	CTCGTGATGCAAGACCGCGCTCGCGGGGCTGTGTGATGTGACGGCATCAAGATCATCAG	633						
QY	691	at 692							
Db	634	AT 635							
RESULT	9								
LOCUS	AW160136	278 bp	mRNA	EST	01-MAY-2000				
DEFINITION	pbkt12 soybean, century cDNA library Glycine max cDNA similar to								
ACCESSION	SEPTUM SITE-DETERMINING PROTEIN, mRNA sequence.								
VERSION	AW160136.1								
KEYWORDS	AW160136.1 GI:7673729								
SOURCE	EST.								
ORGANISM	soybean.								
REFERENCE	Glycine max								
AUTHORS	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.								
TITLE	1 (bases 1 to 278)								
JOURNAL	Matthews,B.F., Devine,T.E., Weisemann,J., Beard,H.B., Lewers,K.S., MacDonald,M.H., Park,Y.B., Maitli,R., Lin,J.J., Kuo,J., Pedroni,M.J., Cregan,P.B. and Saunders,J.A.								
COMMENT	A soybean genetic map incorporating cDNA and EST markers and anchored with SSR markers								
	Unpublished (2000)								
	Contact: Beard HB								
	Soybean and Alfalfa Research Lab								
	USDA ARS PSI								
	10300 Baltimore Ave, Bldg 006 Rm115, Beltsville, MD 20705, USA								
	Tel: 3015045376								
	Fax: 3015045728								
	Email: hbeard@ars.arsusda.gov								
	Insert length: 500 Std error: 0.00								
	High quality sequence stop								

FEATURES	source	location/Qualifiers
		1. 278
		/organism="Glycine max"
		/cultivar="Century"
		/db_xref="taxon:3847"
		/clone_lib="soybean, century cDNA library"
		/note="Vector: pBluescript SK-"
BASE COUNT	64 a	58 c 88 g 67 t 1 others
ORIGIN		
Query Match	16.5%	Score 161.8; DB 40; Length 278;
Best Local Similarity	76.0%	Pred. No. 11e-39;
Matches 199; Conservative	0;	Mismatches 63; Indels 0; Gaps 0;
QY	703	agaagcgaactgatatattaaagagagagatacgaatcagttgtatgacgtgcagag 762
Db	17	AGGGTCGCGACCGACATGATCAAGGGAGGACATGATGTGCGTTGTGACGTGCACAG 76
QY	763	atgtcggagatgcatcttgcttgggtgaatctctgaagatctgaagttatccgaagcag 822
Db	77	ATGTATGGGCTTGGCCCTTCTCGGGTTATTCCTGAGGATAGAGAGTTATTAAGAAGCAC 136
QY	823	aatcgaaggttcgcgccttgcttcgaataaagcctccctacagcttcgaggaattgcgttga 882
Db	137	AATAGAGGTTTCCCTCTGTGCTCAACAAGCCTCCACATTGGCGGAGTTGGCGCTTCGA 196
QY	883	cagcgtgccttggaagcctcgttgagaagaatgatatgaagcgtcttatgtgtgaggaaga 942
Db	197	CNAGCGCGGTGGAGGCTGTGTGAGCAAGATAGCATGACGAGCGCGNGTGTGAGAACAA 256
QY	943	cctaagaacgtgcgctctctct 964
Db	257	CCCAACGTCGGGTTTCTCTCT 278
RESULT 10		
BE124676		
LOCUS	BE124676	362 bp mRNA EST 07-SEP-2000
DEFINITION	EST393711 GVN Medicago truncatula cDNA clone pgvn-67C12, mRNA sequence.	
ACCESSION	BE124676	
VERSION	BE124676.1	GI:8529233
KEYWORDS	EST.	
SOURCE	barrel medic.	
ORGANISM	Medicago truncatula	
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.	
REFERENCE	1 (bases 1 to 362)	
AUTHORS	Fedorova,M., Pearson,B.L., Samac,D.A., Vance,C.P., Gant,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,E.E. and Fraser,C.M.	
TITLE	ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Carroll P. Vance Department of Agronomy and Plant Genetics University of Minnesota 411 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA Tel: 612-625-7219 Fax: 651-649-5058 Email: vance004@maroon.tc.umn.edu University of Minnesota name: M261544e TIGR sequence name: MTCB18TK More information is available at: http://chrysis.tamu.edu/medicago Seq primer: SKmod (CTA gaa cta gctg gat cc).	
FEATURES		
Source		location/Qualifiers
	1. 362	
	/organism="Medicago truncatula"	
	/cultivar="genotype A17"	
	/db_xref="taxon:3880"	







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:

January 8, 2001, 03:36:26 ; Search time 2611.99 Seconds  
(Without alignments)  
1916.225 Million cell updates/sec

Title: US-09-553-431-1  
Sequence: 1 atgagcgtctctgagatgctt.....tcttctcttcttcttgcgcgc 978  
Perfect score: 978

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_cm:\*  
5: gb\_cm:\*  
6: gb\_cm:\*  
7: gb\_cm:\*  
8: gb\_cm:\*  
9: gb\_cm:\*  
10: gb\_cm:\*  
11: gb\_cm:\*  
12: gb\_cm:\*  
13: gb\_cm:\*  
14: gb\_cm:\*  
15: gb\_cm:\*  
16: gb\_cm:\*  
17: gb\_cm:\*  
18: gb\_cm:\*  
19: gb\_cm:\*  
20: gb\_cm:\*  
21: gb\_cm:\*  
22: gb\_cm:\*  
23: gb\_cm:\*  
24: gb\_cm:\*  
25: gb\_cm:\*  
26: gb\_cm:\*  
27: gb\_cm:\*  
28: gb\_cm:\*  
29: gb\_cm:\*  
30: gb\_cm:\*  
31: gb\_cm:\*  
32: gb\_cm:\*  
33: gb\_cm:\*  
34: gb\_cm:\*  
35: gb\_cm:\*  
36: gb\_cm:\*  
37: gb\_cm:\*  
38: gb\_cm:\*  
39: gb\_cm:\*  
40: gb\_cm:\*  
41: gb\_cm:\*  
42: gb\_cm:\*  
43: gb\_cm:\*

44: em\_hcg8:\*  
45: em\_hcg9:\*  
46: em\_hcg10:\*  
47: em\_hcg11:\*  
48: em\_hcg12:\*  
49: em\_hcg13:\*  
50: em\_hcg14:\*  
51: em\_hcg15:\*  
52: em\_hcg16:\*  
53: em\_hcg17:\*  
54: em\_hcg18:\*  
55: em\_hcg19:\*  
56: em\_hcg20:\*  
57: em\_hcg21:\*  
58: em\_hcg22:\*  
59: em\_hcg23:\*  
60: em\_hcg24:\*  
61: em\_hcg25:\*  
62: em\_hcg26:\*  
63: em\_hcg27:\*  
64: em\_hcg28:\*  
65: em\_hcg29:\*  
66: em\_hcg30:\*  
67: em\_hcg31:\*  
68: em\_hcg32:\*  
69: em\_hcg33:\*  
70: em\_hcg34:\*  
71: em\_hcg35:\*  
72: em\_hcg36:\*  
73: em\_hcg37:\*  
74: em\_hcg38:\*  
75: em\_hcg39:\*  
76: em\_hcg40:\*  
77: em\_hcg41:\*  
78: em\_hcg42:\*  
79: em\_hcg43:\*  
80: em\_hcg44:\*  
81: em\_hcg45:\*  
82: em\_hcg46:\*  
83: em\_hcg47:\*  
84: em\_hcg48:\*  
85: em\_hcg49:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	978	100.0	981	AB030278	AB030278 Arabidops
2	978	100.0	65958	AB009056	AB009056 Arabidops
3	452.6	46.3	1164	AF251019	AF251019 Tagetes e
4	247.4	25.3	118360	AF166114	AF166114 Mesostigm
5	243.6	24.9	150613	AB001684	AB001684 Chlorella
6	224.6	23.0	200799	AF137379	AF137379 Neophrosel
7	224.6	22.9	121524	AF041468	AF041468 Gullardi
8	224.2	22.0	22085	PW1245645	PW1245645 Proteobact
9	214.8	19.9	135638	SYCSLRG	SYCSLRG Synecocyst
10	194.2	19.6	194509	AP001129	AP001129 Oryza sat
11	192	13.7	673	AF149810	AF149810 Aquifex a
12	133.6	13.3	14103	AE000712	AE000712 B. subtilis
13	130.4	13.2	1900	BSMTNCD	BSMTNCD Bacillus su
14	128.8	13.2	4389	2	2 BACIREMIN
15	128.8	13.2	5641	2	2 BACIREMIN
16	128.8	13.2	218410	2	2 BSUB0015
17	128.8	12.5	294250	2	2 AP001517
18	122.4	11.2	16486	1	1 AE001931
19	109.4	10.4	8039	1	1 AE001824
20	101.4	9.0	12165	1	1 AE003965
21	87.6				

22	81.4	21607	1	AE001468	AE001468 Helicobac
23	79.8	8.2	13818	1	AE004747 Pseudomon
24	79	8.1	2400	2	ECOM1B
25	79	8.1	11998	1	AE000216 Escherich
26	79	8.1	17045	2	D90751 Escherichia
27	79	8.1	17447	2	D90752 Escherichia
28	78.8	8.1	340806	2	NMA122491
29	76.2	7.8	11184	1	AE004271 Helicobac
30	75	7.7	12907	1	AE000551 Helicobac
31	74	7.6	9966	1	AE002374 Neisseria
32	67.4	6.9	994	77	CNS01KAX
33	65	6.6	293181	2	AP001119 Buchnera
34	57.2	5.8	12730	1	AE005101 Halobacte
35	57	5.8	110071	1	AE000937 Methanoba
36	53.6	5.5	11490	1	AE005032 Halobacte
37	50.6	5.2	10699	29	U67504
38	49.4	5.1	45013	7	F22M8
39	46.6	4.8	7218	81	I66494 Sequence 14
40	46.6	4.8	319000	2	AP000006 Pyrococcu
41	46.2	4.7	9722	1	AE002534 Neisseria
42	44.8	4.6	11813	1	AE001172 Borrelia
43	44.8	4.6	13494	1	AE001056 Archaeogl
44	44.6	4.6	293250	2	CNSPAX02
45	42.8	4.4	12637	1	AE005148 Halobacte

## ALIGNMENTS

RESULT 1  
AB030278 981 bp mRNA PLN 11-FEB-2000  
LOCUS AB030278  
DEFINITION Arabidopsis thaliana mRNA for Mind, complete cds.  
ACCESSION AB030278.1 GI:6759276  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Arabidopsis thaliana

Arabidopsis thaliana cDNA to mRNA.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.

REFERENCE  
AUTHORS Kim, M., Fujiwara, M., Kanamaru, K., Tanaka, K. and Takahashi, H.  
TITLE Arabidopsis thaliana mind homolog involved in plastid division  
JOURNAL Unpublished (1999)  
REFERENCE 2 (bases 1 to 981)  
AUTHORS Kim, M., Fujiwara, M., Kanamaru, K., Tanaka, K. and Takahashi, H.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUL-1999) to the DDBJ/EMBL/GenBank databases. Hideo  
Takahashi, Institute of Molecular and Cellular Biosciences, The  
University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-0032, Japan  
(E-mail: htakahashi@mol.f.u-tokyo.ac.jp/genetics/mgb.html,  
URL: http://imcbs.iam.u-tokyo.ac.jp/)  
Tel: 81-3-5841-7825, Fax: 81-3-5841-8476

FEATURES  
source  
1..981  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="v"  
1..981  
/gene="mind"  
1..981  
/gene="mind"  
/function="mind homolog involved in plastid division"  
/codon\_start=1  
/product="Mind"  
/protein\_id="BA090261.1"  
/db\_xref="GI:6759276"  
/translation="MASLRIFSTNHOSLLPSSLKTKLSSPRFVNPNSRSPISRV  
LOPNRRPELAGETPRIVITSGGKGGKTTTANVGLSLARVGSFVATDADGLRNL  
DILIGENRYNTCEVINGDCRLDQALVDRKWSNELCLISKPSKSLPMGGGKRL  
EMLVDAKTRPESPDITIDCPAGIDAGFITAITPANEAVLVTTPTIALRDADRYT

Query Match 100.0%; Score 978; DB 6; Length 981;  
Best Local Similarity 100.0%; Pred. No. 3.8e-271;  
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT	253 a	205 c	249 g	274 t
ORIGIN	GILECDGIRIDIKMIYVRVPTDMKIGEDMMSVLDVOEMILGLSLGVIPEDESEVIRSTNR GFFVLNKPPTLAGLAFEOAMRIYEDQSMKAVMVEEPKRGFFSFFGG"			
QY	1 atgcgcgtctgagatgttctcaacgaatcatcatcttctccctcatcatctc 60			
DB	1 ATGGCGTCTCGAGATTGTTCTCAACGAATCATCAATCTTCTCTCCATCATCTCTC 60			
QY	61 tcacaaagactcctaatactctcaacgaatctgcaataaccctagaacgagatcca 120			
DB	61 TCACAAAAGACTCTAATATCTTTCACCAAGATTGCGATATACCTAGACGAGATCCA 120			
QY	121 atagatcgtctctcaatctcaatcgaaccggaactcgcgaggaacgcgqatlc 180			
DB	121 ATAGCATCGCTTCTCAATTTAATCGCAACCGGAATCGCGGAGAAACCGCGGTATC 180			
QY	181 gtcgtatcaactccggaagaagcggtgttggaagagacacacacgcaaatgctg 240			
DB	181 GTCGTATCACTCCGGAAGAAGCGGTGTGGAAGAGACACACCAACGCAATGTCGT 240			
QY	241 ctctctcgtcgtctgaagttctcagttctgcgcattgagcgccgaccttgctcgt 300			
DB	241 CTCTCTCGTCTGAGTTTACGTTTCTCAGTTTCTGCGCATTCAGCGACCTTGTCTCGGT 300			
QY	301 aacctcgatcctctcctagagtagagaatcgatcaatctgclcgagagttata 360			
DB	301 AACCTCGATCTCTCCTAGAGTAGAGAAATCGAATCGAATTAACCTTGCCTCGAGTTATA 360			
QY	361 aacgagatgtcgtctcgatcaacagctcgtgtaaggaaggttgctgcgaattcgaa 420			
DB	361 AACGAGATGTGCTGTCGATCAAGCTCTGTCGTAATGAGCTTGCTGCGAATTCGAA 420			
QY	421 ttgctatgatacttaac 480			
DB	421 TTGCTATGATATCTTAAC 480			
QY	481 gaatgcttgttgatcgttgaaacacacacacacacacacacacacacacacacacac 540			
DB	481 GAATGCTGTGATCGCTTGAAACACACACACACACACACACACACACACACACACACAC 540			
QY	541 gattgctcgtcgaagatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgat 600			
DB	541 GATTGCTCTCGAAGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 600			
QY	601 gtctgtaaac 660			
DB	601 GTCTGTAAAC 660			
QY	661 ttagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtag 720			
DB	661 TTAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 720			
QY	721 attaaagagagatagatgtagatgtagatgtagatgtagatgtagatgtagatgtag 780			
DB	721 ATTAAGAGAGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780			
QY	781 ctgtgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 840			
DB	781 CTGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 840			
QY	841 gtctgtaaac 900			
DB	841 GTCTGTAAAC 900			
QY	901 gttgagcaagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtag 960			
DB	901 GTTGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960			

QY 961 ttctctctcttgccgagc 978  
|||||  
Db 961 TTCTCTTCTTTGGCGC 978

## RESULT 2

AB009056 65958 bp DNA 09-AUG-2000  
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MZF18.  
DEFINITION  
AB009056  
VERSION  
KEYWORDS  
SOURCE

## ORGANISM

Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1  
clone:MZF18  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsi.

## REFERENCE

## AUTHORS

Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N.  
and Tabata,S.

## TITLE

Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
nineteen physically assigned P1 and YAC clones

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CDS

On Aug 9, 2000 this sequence version replaced gi:2656032.  
Address for correspondence: kaoskazusa.or.jp  
For the latest information on annotation of this clone, please see  
http://www.kazusa.or.jp/kaos/cgi-bin/seqgraph.cgi?c=MZF18  
Genes with similarity to proteins in the databases are described in  
'product' or 'note' qualifiers. Genes that have no significant  
protein similarity are described as 'unknown protein'.  
The software programs used to predict genes include: Grail  
(Informatics Group, Oak Ridge National Laboratory,  
http://compbio.ornl.gov/Grail-1.3/),  
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of  
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
SplicePredictor (Volker Brendel, Stanford University,  
http://gremli.zool.jastat.ehu/cgi-bin/sp.cgi).  
Genes encoding tRNAs are predicted by tRNAscan-SE  
(Sean Eddy, Washington University School of Medicine, St. Louis,  
http://genome.wustl.edu/eddy/tRNAscan-SE/).  
This sequence may not be the entire insert of this clone. It may be  
shorter because we remove overlaps between neighboring submissions.  
The 5' clone is M001 and the 3' clone is M185.  
Location/Qualifiers  
1. 65958  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="MZF18"  
clone\_lib="Mitsui P1"  
join(355..1064,1116..5684)  
/note="gene\_id:MZF18.17"  
/codon\_start=1  
/pseudo  
/evidence=not\_experimental  
/product="non-LTR retroelement reverse transcriptase-like  
protein"  
join(8240..8326,8416..8695,8789..9161,9246..9406,  
9571..9698,9776..10024,10109..10399)

## FEATURES

## source

## CDS

1. 65958  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="MZF18"  
clone\_lib="Mitsui P1"  
join(355..1064,1116..5684)  
/note="gene\_id:MZF18.17"  
/codon\_start=1  
/pseudo  
/evidence=not\_experimental  
/product="non-LTR retroelement reverse transcriptase-like  
protein"  
join(8240..8326,8416..8695,8789..9161,9246..9406,  
9571..9698,9776..10024,10109..10399)

/note="gene\_id:MZF18.16"  
/codon\_start=1  
/evidence=not\_experimental  
/product="delta-cadherine synthase (d-cadherine  
synthase)"  
protein\_id="BAB08719.1"  
/db\_xref="GI:9758220"  
/translation="MGSEVNRPLADPANIIMEDPLTFSKSDLGTEFTFKKSTLKEA  
VKEAFSSKANPIENIKFIDALCRGVSVHFEKDIYEQDLDKSDCLDFQWNRQECDD  
LYTVGIIIOVDFPQFGPKLSADVPKFRKNGFKGLHLYDAGMSLDFEQAQMGTHGE  
CDIIDLAFNRSHLEISRSRSPHATIRKLNKLPYHNGISRIETROYISYEES  
PEPQSQARVITTMALILPTALDDIDDEIPIDSMKFTYVNVVDFRIRITAYLMSCTY  
CGGHLKSLQKTANGVQOAKWMLKDVIAPEDEYKNAIISSGYVALDEEIEKGRS  
VAKDLAEFMLSHPKIRVASEKISRSFTDISEFEKREHVAATGIDCQMOGVSK  
KEHIVSLIETIP1"  
11597..12883  
/note="gene\_id:MZF18.15"

## CDS

/evidence=not\_experimental  
/product="acetyl-CoA:benzylalcohol acetyltransferase-like  
protein"  
protein\_id="BAB08720.1"  
/db\_xref="GI:9758221"  
/translation="MKLELSEKVKPASPNIQTLSTLFDQPLPTVYVAFIPEYD  
HSNODIORLKSLSQTLSTLFLYPLAGQIKDQVHCNDEGLFETARKEITLSDTLN  
PSDADLQKFTVSPDHADPEWPLHYKVFPPKGRVAVASVSKICDANSLSTFC  
SMTRASKGVADTVDPPEVGNDFPPADISIEPPLVHETSKTRKREYVGSIMKELK  
HKAICNFPFLDLKSSBSKMEIEVYSKQKQGVKVEFAITQIMDLRPVSSSLP  
ASAMLSLSEIPMETEYVAVSSWCMRSMSEYATPEPSTLI"  
complement(join(13438..13580,13695..14419,14516..14753,  
14956..15283,15400..15634,15951..16167,16240..16311,  
16400..16553))  
/note="gene\_id:MZF18.14"

## CDS

/evidence=not\_experimental  
/product="FRO2 homolog"  
protein\_id="BAB08721.1"  
/db\_xref="GI:9758222"  
/translation="MGNMRSLVKTLMVFLGLWLMVIMINISTNLFKSKMPTSKYIN  
TYVEGQGNVLVLTVPMMFLVAVLSGYVILIRKLSQSYIECEKMKLRMGVIT  
MNPGLVITATELFTSLFVALMLASNYLYSLHYHLHNDNAKIQOAFRAGKVI  
GYGNICMAFLFPPVTRASTIILPLVGLTSSTSKYHMGICHSVECEGLHTVFLIIV  
AMINKMEFEANPVTVPNLAGTIAWVIGATKMTSPSTRKKFELEFTTHLYGLY  
IYEVYIHVGDMSWCMVTPNLEFTFIDRIYLFQDSTKRSLSAAILPSDNLELFSKT  
SGIDSLSEVTECPYSGNSFDVSRHNSLIVSGSGITPFTISVIRRODSWOKLYTHLS  
DYLVCSERKHYHDLAFLDLIFPLDMSASDISRLNLEAVITREDDKPEPTDHRILQ  
TWKFKQPLDPSISPVLPNNFLMLGVILSVLSEVMEFLIGITRYIYVQHNIGSI  
YNEFSYGLMDMEFGACIFISSSVFLMRKQKQKEDKERNVOVSEVQPTSPRS  
WPHGERLEKLESPVOSIYOATSVHFGSKPLKILLEAESDEVCWVCCPKRRMRHEV  
AKICSSGILSLHPEAISFNW"  
complement(join(20836..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.13"

## CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="FRO2 homolog"  
protein\_id="BAB08722.1"  
/db\_xref="GI:9758223"  
/translation="MGNMRSLVKTLMVFLGLWLMVIMINISTNRONTITPKLAKYLK  
TYVEGQGNVLVLTVPMMFLVAVLSGYVILIRKLSQSYIECEKMKLRMGVIT  
MNPGLVITATELFTSLFVALMLASNYLYSLHYHLHNDNAKIQOAFRAGKVI  
GYGNICMAFLFPPVTRASTIILPLVGLTSSTSKYHMGICHSVECEGLHTVFLIIV  
AMINKMEFEANPVTVPNLAGTIAWVIGATKMTSPSTRKKFELEFTTHLYGLY  
IYEVYIHVGDMSWCMVTPNLEFTFIDRIYLFQDSTKRSLSAAILPSDNLELFSKT  
SGIDSLSEVTECPYSGNSFDVSRHNSLIVSGSGITPFTISVIRRODSWOKLYTHLS  
DYLVCSERKHYHDLAFLDLIFPLDMSASDISRLNLEAVITREDDKPEPTDHRILQ  
TWKFKQPLDPSISPVLPNNFLMLGVILSVLSEVMEFLIGITRYIYVQHNIGSI  
YNEFSYGLMDMEFGACIFISSSVFLMRKQKQKEDKERNVOVSEVQPTSPRS  
WPHGERLEKLESPVOSIYOATSVHFGSKPLKILLEAESDEVCWVCCPKRRMRHEV  
AKICSSGILSLHPEAISFNW"  
complement(join(20836..20978,21097..21821,21910..22147,  
22283..22610,22727..22973,23265..23456,23532..23628,  
23708..23861))  
/note="gene\_id:MZF18.13"

CDS

KLQTKWFKPDLSPISPLGPNLMLGVLLISSVWFLLLIGIVTRYYPVBN  
TGIYNTYRYLWVWMLGVCIPFISSSIIILMKREKEDKSKKVOSEFPTPS  
SPSGMHEHELESYVOSIVQATSVHFGSKPLMKLLEAESESEVGVWVCKPKM  
RHEVAKTCSGGLAKNLHFEAISEFM"  
complement (join(25651..25974,26145..26386,26718..26876,  
27106..27216,27399..27625,27695..27963))  
/note="contains similarity to unknown protein  
dbj|BA90637.1  
gene\_id="MF18.12"

CDS

/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB08723.1"  
/db\_xref="GI:9758224"  
/translation="MALFIVVSPNPISPIISSIHYSPSRHLSSTITIRNKPSPV  
LTPIRLVSSRRRYKVSQLEPLISPDHMSQWALAEAGAFVSEKRTIGSKWSP  
ALSTLGLAASNLRIPEFEPYGFEEFELPHTIPLELRADLRIRISTGSLLA  
PLIGSVATIGTVAAFMVPMRSLGPDMMKIAALMGSYIGISINFAVIAISALISPS  
VIAGAVANDVICALHEMVLPALASKIPPEASASPADMTKDKLEDKNRVSTSI  
ALVSFLICKRAITLTTIFKTOGVMLPAVATITTYLATSPDEFSLAPSETTSLI  
MOVEFTLIGATGVWNVNTAPSLPLPAIDVWMLATVTLGKIFCDMKLLILASN  
ANIGPTTACAMATAKMTSLVPGILSGVGIATPILGICGVFLKRL"  
29181..31655  
/note="gene\_id="MF18.11"  
/codon\_start=1  
/evidence=not\_experimental  
/product="receptor-protein kinase-like protein"  
/protein\_id="BAB08724.1"  
/db\_xref="GI:9758225"

CDS

/translation="MAFPIINQTLLEFFCPDLILSFAAFTPTDYLINSQNTNS  
FTTRSLDSSESGSFLSTDRSISIDNPDPDVLVNTARVPGVGYFQVYTK  
GTHFRLHAPKAPKAPNLRSAPKPVLLNGFVINSFVSIVVVEELIKIDPVTLEI  
SPLPKASGFGVNAVEVSAKDYIMDQGTLYVINSQIENSISQVLEFVHRIN  
GSKITPRNDITMPTVVDNLLRAARAMTHSPYONGAGATREIAPDNVMTA  
CEMDNDQELQARENTSMGFQYDEKRVILVPLHCDIVSSILNOLYEVFVNEYLA  
KDVDLSTLFELHASPILYIDEVAESDRSGMLRISVPSLSPARNALLNVEIMRI  
LSPVSEVSGKRNWVIVVSVLGFEVFLSLFVSLCLCKRNKTRKSTGCTMP  
LRRFGSSNRRTERTVSSGYHTLISPAELQSGNPNDRSLVIGVGFGVFRSL  
KNTVAVVRSRSGPSRQGLPEFLEITILSKIRHRLVSLVGYCEQOSMLVVEYMD  
KPLKSHLYGNTNPNLSMKRQLEVCAGRLHYLHTGSSQGIIRHDIKSTNILLDN  
YVAKVADPFLSSRQCIIDETHTSTGVKSGFYLDPYFRROOLDKCSLKFETAEK  
VLCARPADVPLREOVNLAENAIEMORGMDOIDVPMIADETIKCSLKFETAEK  
CCADYGVDPRTIGDVLMNLEHVLQLOESQPLNPEIDYDVIDPRARQGLSGNSIE  
RDYGDTSGLISSTOVFSQMLTMAGR"  
32002..32982  
/note="dbj|BA90261.1  
gene\_id="MF18.10"  
/codon\_start=1  
/evidence=not\_experimental  
/product="septum site-determining MIND"  
/protein\_id="GI:9758226"  
/db\_xref="GI:9758226"

CDS

/translation="MASLIRLESTNHOSLLPSLSQKTLISPRFVNPSPRSPIRSV  
/LOPNRPELAGETPRIVITSGKGVGKTTTANVGSILARSFVSVAIDADGLRL  
DLILGLENRYNTCEVYINDCRLDOALVDDTRKSNSEILICISKPSKLMADGKRL  
EMLVADLKTREPSPDITIDPAGIDAGTTRITPANNEVILVTPDITLALRADRYT  
GLEDCGIRDIKIMVKNRTMDIKGEDMSVLDVDEHGLISLGLVPEDEVIRSTRK  
GPELVLPPTLGLAFEEQAARLVLEDSDSKAAVVEEPRKRFESFVG  
complement (join(33889..34276,34361..34392,34482..35252,  
35439..35984,36094..36264))  
/note="contains similarity to unknown protein  
gene\_id="MF18.9  
pir|I702898"

Query Match 100.0%; Score 978; DB 6; Length 65958;  
Best Local Similarity 100.0%; Pred. No. 6,9e-271;  
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcgtctcgtgagatgttccaagaatacaatctctctcctccatcatctctc 60  
|||||  
Db 32002 ATGGCGTCTGAGATGTCTTCAACGAAATCATCAATCTCTTCCATCATCTCTC 32061  
|||||  
QY 61 tcacaaagacttaattctcaccagatcgtcacaaccctagacagagatcca 120  
|||||

Db 32062 TCACAAAGACTTAATATCTTCAACGAAGATTGCTAATAACCTTAGCAGACGAGTCCA 32121  
QY 121 ataagatccgtcttccaatttaattcgaacacggaaacccgcggaagaacccgcgaatc 180  
|||||  
Db 32122 ATACATCCGTCCTTCATTTAATGCAACCGAACTCGCGGAGAAACCGCGGTATTC 32181  
|||||  
QY 181 gtcgttacacccctccggaagaagcggtgttggaagaagacaaccaacgcgaatgtcggtc 240  
|||||  
Db 32182 GTCGTATCACCTCCGGAAGAGCGGTGTGGAAGAGACACACCGCAATGTCTCGT 32241  
|||||  
QY 241 ctctctcgtcgttgaaggtttctcagtttgcgcatttgagcgcgacttggtctccgt 300  
|||||  
Db 32242 CTCCTCTCCGCTCTTACGCTTTCAGTGTGCGCATTTGAGCGCGACTTGTCTCCGT 32301  
|||||  
QY 301 aaccctcgtctccctcctaaggggttaagaaatcgaatcaacttgcgtcgaagtata 360  
|||||  
Db 32302 AACCTCGATCTCTCTCTAGAGTTAAGAAATCAGTCAATATACATCTGCGTAGGTTATA 32361  
|||||  
QY 361 aacgagatgtcgtctcgaatcaagctcgtgaagtataagcgttgcgaatttgaa 420  
|||||  
Db 32362 AACGGAGTTGTCTGCTCGATCAAGCTCTGTACGATACGATGATGATGATGATGATGAT 32421  
|||||  
QY 421 ttgctatgtatcttaaaccttagatcgaactcgaatggaatttggtgtaagcatgt 480  
|||||  
Db 32422 TTGCTATGTATATCTTAACCTAGATCAATCCATGCGATGCGATGCGATGCGATGCGAT 32481  
|||||  
QY 481 gaatgctgttgatgtcgttgaaactgaacggaaaggttcacggatctcatcatc 540  
|||||  
Db 32482 GAATGCTGTGTGATGCTGTAAGAAACTAGACCGGAAAGGTTCACCGGATTCATCATCATC 32541  
|||||  
QY 541 gattgtccgcggaatcgtatgcggatcacaacacccatcctccgcggaatgaaga 600  
|||||  
Db 32542 GATTGTCTGCAAGAAATCGATGCGGATTCATATACCGCATTTACTCCGGGATGAAAGCA 32601  
|||||  
QY 601 gttctgtaacaactccggaatataaacgcgttlaaggaatgtatgaggtttacgggtttg 660  
|||||  
Db 32602 GTTCTGTATCAACTCCGATATACAGCGTAAAGGATCTGATAGCGTTCACCGGTTTG 32661  
|||||  
QY 661 ttgaatcgatggaatcagaatataaagatgttgaaaggaaggaaggaatcgaatg 720  
|||||  
Db 32662 TTGAATCGATGGAATCAAGATATTAAGATATTTGTGACAGATGAACTGATNTG 32721  
|||||  
QY 721 attaaagagagatatgacgtcaggttgatgttgacaggaatgttgaggttgatcgtg 780  
|||||  
Db 32722 ATTAAAGAGAGATATGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 32781  
|||||  
QY 781 ctgtgttaattccctgaagatctcgaatgtatctgaagacgaatcgaaggtttccggt 840  
|||||  
Db 32782 CTGTGTAAATTCCTGAAGATTCGAGGTTAATGGAAGACAGATGAGGTTTCCGCTT 32841  
|||||  
QY 841 gttctgaataagcctcctcagctcgtcggaatgtgcttgcgagcgagcttgaagatc 900  
|||||  
Db 32842 GTTCTGAATTAAGCCTCTTACCTTCCGAGATGTGCGGCTTTGAGCAGCGGCTTGAAGATC 32901  
|||||  
QY 901 gtcgacagaatgatgatgaagcgttatagttggaagaagaacctgaagaagcgtgcttc 960  
|||||  
Db 32902 GTTAGCAAGATATGATGAAGCGTGTATGTGTGAGGAAGAACTAAGAAACGTGGCTTC 32961  
|||||  
QY 961 ttctcttcttcttgagcgcg 978  
|||||  
Db 32962 TTCTCTTCTTCTTGGCGCG 32979  
|||||

RESULT 3  
AF251019 AF251019 1164 bp mRNA PLN 06-SEP-2000  
LOCUS Tagetes erecta mind mRNA, complete cds.  
DEFINITION Tagetes erecta mind mRNA, complete cds.  
ACCESSION AF251019  
VERSION AF251019.1 GI:9971821  
KEYWORDS  
SOURCE African marigold.  
ORGANISM  
Tagetes erecta  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;







MAHNPDLDAVEAPAVNG"  
6178..6300  
/note="ORF40a"  
/codon\_start=1  
/protein\_id="BAA57843.1"  
/db\_xref="GI:2224359"  
/translation="MVRCCSHHETFSNKRCEVLCVQLQNTLYKKAVYELRDS"  
complement(6212..6379)  
/note="ORF55a"  
/codon\_start=1  
/protein\_id="BAA57844.1"  
/db\_xref="GI:2224360"  
/translation="MLAVKATGLBERNDLRTQIQOLGVLPNTLSAHKQPPYIEFAG  
HILHIAVY"  
6803..6943  
/note="ORF46a"  
/codon\_start=1  
/protein\_id="BAA57845.1"  
/db\_xref="GI:2224361"  
/translation="MESKQISFFIFFSFFFLFFFIKKERASPAFIPLSFLFCFYK  
SR"  
7448..7693  
/gene="psac"  
7448..7693  
/gene="psac"  
/codon\_start=1  
/product="Photosystem I iron-sulfur center"  
/protein\_id="BAA57846.1"  
/db\_xref="GI:2224362"  
/translation="MSHVKIYDTCIGCTGCACPTDVLKVPMDGKASQIASAPR  
TEDVGCRCESACPTDPLSVRVYLGSETTRSMGLAY"  
7840..7989  
/note="ORF49a"  
/codon\_start=1  
/protein\_id="BAA57847.1"  
/db\_xref="GI:2224363"  
/translation="MKELTFKFFRIASFELRKILFQSOKSKIIISWRIKRIITFEF  
LECDs"  
complement(7986..8297)  
/note="ORF103"  
/codon\_start=1  
/protein\_id="BAA57848.1"  
/db\_xref="GI:2224364"  
/translation="MNVFDLILITFTLYGNLFAIQNCVLMHNPVFIYIATVEFN  
KIIFYLQKKKDVQSRVFEFTKKNDVMPNPFVLSNIIITVKRGFLGLFVEAEKVG  
S"  
8330..8413  
/gene="trnI"  
/product="tRNA-Ile(CAU)"  
8330..8413  
/gene="trnI"  
complement(9037..9183)  
/note="ORF48a"  
/codon\_start=1  
/protein\_id="BAA57849.1"  
/db\_xref="GI:2224365"  
/translation="MFSKKKFSKYFILNLYDPTSSKKMMFCLQFGFILRLIAKNFN  
KTNK"  
complement(9141..9557)  
/note="ORF138"  
/codon\_start=1  
/protein\_id="BAA57850.1"  
/db\_xref="GI:2224366"  
/translation="MENATVPNDONEDYIDLTKPMKSYWKKMKVKGFRWLVLDK  
KINDILNTEIYPIIVSGAKFTKSRFSAMRSLNFAFVDDGFLLPFFYKKKRS  
KISNYSKKRRKKRTRNLGNLCFLRNKLASTLFF"  
9485..9745  
/note="ORF86b"  
/codon\_start=1  
/protein\_id="BAA57851.1"  
/db\_xref="GI:2224367"  
/translation="MTYSYLNLNCLNLRGHLERLHPSETTIOMTFHKKRKLFLFL  
FEFCQERSVSIKIRPKTNHTGVARLLVKKRIILTKAKETLHY"

complement(9762..10163)  
/note="ORF133"  
/codon\_start=1  
/protein\_id="BAA57852.1"  
/db\_xref="GI:2224368"  
/translation="MVTLSDESNAASSVTKNEVNTSSLVNGVLLITENVNRKHSV  
ENQASSTEEQTLKISKSNINPQKSSPFLGNTSPVIRPTSVFKSTVLSTENNL  
VSFEVDFMNFESALNLFLESILYEVNDFS"  
10753..10875  
/note="ORF40b"  
/codon\_start=1  
/protein\_id="BAA57853.1"  
/db\_xref="GI:2224369"  
/translation="MEARKSNLSDPLNFGNSRRKRGVHNKRRKGGKATNVL"  
complement(11039..11203)  
/note="ORF54a"  
/codon\_start=1  
/protein\_id="BAA57854.1"  
/db\_xref="GI:2224370"  
/translation="MSVNLTEKNDQKLFVSOEIIITGVIALFCCVEMGPTLELSFYG  
FSYVLGYLS"  
complement(11907..12113)  
/note="ORF68"  
/codon\_start=1  
/protein\_id="BAA57855.1"  
/db\_xref="GI:2224371"  
/translation="MFSNLGKFLIKSVTKVINKHITSNLELTKKRGOLFRRHREN  
RISFGSOWLIYGNKMKIFLFLFF"  
complement(12455..13975)  
/gene="atpA"  
complement(12455..13975)  
/gene="atpA"  
/codon\_start=1  
/product="ATP synthase CF1 alpha chain"  
/protein\_id="BAA57856.1"  
/db\_xref="GI:2224372"  
/translation="MVKIRPDEISSIIKQITROYQOEKAVNVGTVPQVGDRIATYG  
LDKVMAGLVEFEDGTGIALNLKKNVGAVALMGSTKRYOEGSSYRATKIQIYED  
GYLGRVNSLARPIDQGEIATKENRLIESPAPGIISRSVHEPIQTGTVAIDAMIP1

Query Match 24.9% Score 243.6; DB 6; Length 150613;  
Best Local Similarity 59.5%; Pred. No. 66-59;  
Matches 439; Conservative 0; Mismatches 284; Indels 15; Gaps 1;

QY 174 gcgtatcgtctatcaccctccggaagcgggtgttggaagacgacacgcgcaaa 233  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 99337 GCCTGAATGTGTATTAACCTACAGCAAGGTGGGTTGGGAAAAACAACAACAGCAAA 99278  
QY 234 tctcgtctctctcgtcgtctcgttaaggtttctcgaagttgcgaatgcgacgttg 293  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 99277 TTTAGGAATGTCATTTGCTGGCTTGCTGCTGCTTATTTGATGCAGATATAGG 99218  
QY 294 tctcgttaacctgattcctcctcctaggttaggaatgcgaatgcgaatgcgtgcga 353  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 99217 ATTACGTAATTTAGATTATTTATTTAGGATTTGGAATAATCGTACTATCTGCATGGA 99158  
QY 354 ggttaaaacgagatgtctcgtcgtcgaagctcgtgtaagtgatagcgttggtcgaa 413  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 99157 TATTTGTAAGGTCATGATGACATCAAGCTTTATTCGATTAAGATGAAGATGGA AAA 99098  
QY 414 ttctcgaattgtatataactaaactgaatgcgaatgcgaatgcgttggtgtaa 473  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 99097 TTTAGCATTTATTGACATTTCTAATAACACACACAAAA-----TATATATG 99053  
QY 474 agcatggaatgcttgtagatgcgttgaanaactagaacggaaggttacccgattcat 533  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 99052 TACAGAAGAAACATGACGATTTAATTTGATTTCTGTAAGAAGACTCGTTTTCAGTTTGT 98993  
QY 534 catcatcgtatgtcctgcggaagatgcgtccggaattataacgcgcatctactccgcgcaa 593  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 98992 TTTAATTGATTTGCTCTCTGCTGATTTGACGTAAAGCATTTGATTCATCTGCACA 98933  
QY 594 tgaagcagttcttgtaaacactccggaatataacagcgttgaaggatgctgaaggttac 653

gene  
tRNA

/gene="trnL(uua)"  
/note="codon recognized: UUA"  
/product="tRNA-Leu"  
/anticodon=(pos:798..800,aa:Leu)  
765..849  
/gene="trnQ(uaa)"  
983..1054  
/gene="trnC(uug)"  
/note="codons recognized: CAU"  
/product="tRNA-Gln"  
/anticodon=(pos:1015..1017,aa:Gln)  
983..1054  
/gene="trnO(uug)"  
1102..1182

gene  
tRNA

/gene="trnI(uaa)"  
/note="codons recognized: CUG"  
/product="tRNA-Leu"  
/anticodon=(pos:1136..1138,aa:Leu)  
1102..1182  
/gene="trnII(uaa)"  
complement(1364..1468)  
/gene="psbM"  
complement(1364..1468)  
/gene="psbM"

CDS

/codon\_start=1  
/product="P protein of photosystem II"  
/protein\_id="AAD54779.1"  
/db\_xref="GI:5880686"  
/translation="MEVNIQLIALFLIPTSFLLIVKTAQQN"  
1618..1691  
/gene="trm(cau)"  
/note="codon recognized: AUG"  
/product="tRNA-Met"  
/anticodon=(pos:1652..1654,aa:Met)  
1618..1691  
/gene="trm(cau)"  
1733..3862  
/gene="ftsI"  
/gene="ftsI"  
1733..3862  
/gene="ftsI"

gene  
CDS

/codon\_start=1  
/product="putative peptidoglycan synthetase"  
/protein\_id="A054780.1"  
/db\_xref="GI:5880687"  
/translation="MIYPMKYRFRRLPKHLGGISYYAFISTMTIKILADPTSKPKMHSGTFEYNERNRWIKSLATVCLEFTRLSGLGFENFKYEKLSSVAERDQIGLVNDPOQRILLDHNGDIVALIDPAYDLVLYHPMCSTLERIELSPLLDSDSERQDIGRVBOSGCITLMHGDINDINTPAOIIRLGVDGICGLVHHPORAVIPKRSPESILGVYTEGRYCAGGLESIIDDMKSTYODVPDWMDGNHNFGIINPRKOILFHOBASLOILTDCGLOLEKEPESTRKPVALAIETEAHSGLRICLATSPSYDPNCYGMFPMEFFRCMPITLTGTQIOLRNSNVNRWVSOLDAPIAYIHRLTGLGSHRNNOTSFKMSHDHESEGNLNKDRTSIDAISVVAKDDDEFVDHEIEATAVGCGLAMTPKLILTIATTANGMAVTPLHTSLVIDLFHFHHNLSONNEFSLOGWGWSVLSRSNOVAROPRYTHDIYLGHVPVPSLELGEFTVLDFRHFNLOSMMNFSLQGWSGOVHVLAQAOTSGCELPGAAMPVKQTAAKAALSOGYSTSVYSVFVCIYAAYVRKFVTVII DEPEDPFREFNTADVTSKTLSIMTVQEODDEPPSYPVSLFERNM"  
3398..5059  
/gene="psba"  
3398..5059  
/gene="psba"  
/gene="psba"  
/gene="psba"  
/codon\_start=1  
/product="Dl reaction-center protein of photosystem II"  
/protein\_id="AAD54781.1"  
/db\_xref="GI:5880688"  
/translation="MTAILRRRSTSWARFCMWSTENRIYIGNCVMLPLLTLTDENVLIQCCTVIAPWDVICIGREPYSGSLIFGNNTISSALIIPSSAAIGIHFYPIPEASTVFIQSQSDSNEMPGIGSTEFMFMYEQAEINIIMIPFMIMGVAGVGGSLFSAHGSLAMRPVACIMETFAIGALGYSTMAFNMGINFENOSVDSQGRVINTMDAITRNANLGEVMHERNAHNPIDLASSVDAVAQG"  
5489..6964



## ORGANISM

Chloroplast Nephroselmis olivacea  
Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;  
Chlorodendraceae; Chlorodendraceae; Nephroselmis.

## REFERENCE

1 (bases 1 to 200799)  
Turmel, M., Ollis, C. and Lemieux, C.

## AUTHORS

The complete chloroplast DNA sequence of the green alga  
Nephroselmis olivacea: Insights into the architecture of ancestral  
chloroplast genomes

Proc. Natl. Acad. Sci. U.S.A. 96 (18), 10248-10253 (1999)

## JOURNAL

2 (bases 1 to 200799)  
Turmel, M., Ollis, C. and Lemieux, C.

## AUTHORS

Submitted (24-MAR-1999) Biochimie, Universite Laval, Pavillon  
Charles-Eugene Marchand, Quebec G1K 7P4, Canada

## TITLE

Location/Qualifiers

## FEATURES

## source

1..200799

/organism="Nephroselmis olivacea"

/organelle="plastid:chloroplast"

/db\_xref="taxon:31312"

/cell\_line="NIES 484"

1..92126

/note="Region: large single-copy region (LSC)"

## misc\_feature

55..137

/gene="trnL(gag)"

/note="codons recognized: CUY"

/product="trnL-leu"

/anticodon="(pos:88..90,aa:Leu)"

55..137

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

332..634

/gene="trnL(gag)"

/product="trnL-Met"  
/anticodon="(pos:1652..1654,aa:Met)"  
1618..1691  
/gene="trnM(cau)"  
1733..3862  
/gene="trnI"  
1733..3862  
/gene="trnI"  
/codon\_start=1  
/product="putative peptidoglycan synthetase"  
/protein\_id="AAD54780.1"  
/db\_xref="GI:5880687"  
/translation="MTNPKYRROFRRLPKHLOGIYVAFLSISTMIKIALDPYSKR  
PKMHSSTPFOYENEMRWIKSLATVGLTFATRLSGLOFNKYTELKSVARQIO  
VNDPOQRKIIIDHNGDIVAIDPAVDLYVHPRMCTSLERIAELSPILDSOYL  
RDEGDSGICLMDHIDNTSAQIRLGVDSIEVHHPRORYPKRSPSTLGYDTGE  
YQAGLESSLDMWMSYODVPCMDGHCNFGILRPFKOILPHQESMLQITDCGLOE  
KVSOLITNANMRFGAKRIAIMEHSGATRCATSPSTPNCTGMFPMRFRCPWIT  
DLEPSTFRPVLAIALENGIFOPDRIIDTCKIRIGDSWIGNVGGFIMDSLDHL  
TGOILDRSSNVMVWMSIDPAIYHRNLRIGLSDHNDNOTSRMSSHDHNSGM  
NKLIDSDYAIYVKQDERVDEIEAATASFGQIAMPPLKLILOLIATANGMAAT  
PRLIKTYITLDHFRHLOSNNESLQGVOSYLSRQYAKQRPYTHDLYLGHVVP  
SLELQNFQVKSIPPHTRERRRLESQYCNVLMGLQVVLDAQATSGRLPGYAMG  
KTGNOKASALGYSYDVSIVTSFVGIYPAVKPFYVTLVIDEDPDRFGFNADVIT  
OTLISEMIVQEDPPSYPTVSLFERNM"  
3998..5059  
/gene="psbA"  
3998..5059  
/gene="psbA"  
/codon\_start=1  
/product="D1 reaction-center protein of photosystem II"  
/protein\_id="AAD54781.1"  
/db\_xref="GI:5880688"  
/translation="WTALERRRSTSVARFCDWYTFENRLYIGWGLMIPILITA  
TSVFITGIAAPVDIDGIREVSGSLFNGNIIIGALIPSSAAGIHYRPIWEAS  
DEWLYNGGCYELIVLHFLGVACYGWRMELSEFLGMWPIVAVASAPAAATATFLI  
YPIQGSFSDMPILGISTFNEMIVFOALNIHNPMLHMLGCGSLFSAMHSL  
VTSSILRETTENESANAGYKFGQEEETNYIAAHFEGRLITQVYAFNSRSLHFLA  
AMPVCIWFTALGVSTMAFNGLFNFQSVYDSDGRVINTMADITINRANLGMVHHR  
NAHNPDLASVDPAVAGC"  
5489..6964  
/gene="atpB"  
5489..6964  
/gene="atpB"  
/codon\_start=1  
/product="CF1 beta subunit of ATP synthase"  
/protein\_id="AAD54782.1"  
/db\_xref="GI:5880689"  
/translation="MSQAGVTKQAGTIVGVGMVAVAPGSGMPNITVALYGRN  
GAGQEVNVCVEOOLIGDGLVRAVSMKNTDGLMGEVTDGRALSVPGPTTLGRIF  
NLVGEVNDNMGPVNGEKTPIHREAPAFVVDIDTKLSIFETGIKVVDILAPYRGGKIG  
LFGAGACGVTLIMELNNTAKAHGVSFVGQVEERTREGNDLYMEKESKVIKNADY  
KSKKVALYVQKMEPPGARMKVGLTALTMAYEFDVANDODLLEIDNIFRVOAGSEV  
SALGDMPASVGYQPTPLATEMGGLQESTKDSITSIAVYVAPADLDTPAPATPF  
AHLDATVLSRNLAAGIYPAVDPLDSTMLDPMWILGDHGYGTAORVKOTLQRYKEL  
ODIATLIGDELSEEDRMIVARAKIERFLSOPFEVFTGAGKGYPLAESIQGER  
LTLISGLDSLPEESAFYLVGNIFEAIKAASIQAAAAAAA"  
6977..7378  
/gene="atpE"  
6977..7378  
/gene="atpE"  
/codon\_start=1  
/product="CF1 epsilon subunit of ATP synthase"  
/protein\_id="AAD54783.1"  
/db\_xref="GI:5880690"  
/translation="MSLQVRIITPEQVFLNVSADFIILPTNGMGVLTNHPPLITAI  
DIGPMIVRSESTWQSMALLGLALVLDONOVITLVNABELGSDINABEATETFLAKAEA  
LANSKTRKDOIENNLAFKRARVYQVATLVK"  
complement(7622..8104)  
/gene="pepD"  
complement(7622..8104)  
/gene="pepD"  
/codon\_start=1



FEATURES	COMMENT
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Cryptomonas phi does not involve a signal peptide Plant Mol. Biol. 15 (4), 585-592 (1990) 9138697 6 (bases 110917 to 113854) Douglas,S.E. 1991
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Unusual organization of a ribosomal protein operon in the plastid genome of Cryptomonas phi: evolutionary considerations Curr. Genet. 19 (4), 289-294 (1991) 9133043 7 (bases 40675 to 42376) Douglas,S.E. and Turner,S. 1991 Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor J. Mol. Evol. 33 (3), 267-273 (1991) 92099311 8 (bases 96129 to 98906) Wang,S.L. and Liu,X.Q. 1991
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	The plastid genome of Cryptomonas phi encodes an hsp70-like protein, a histone-like protein, and an acyl carrier protein Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991) 92073372 9 (bases 106789 to 108216) Douglas,S.E. 1991
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	A secY homologue is found in the plastid genome of Cryptomonas phi FEBS Lett. 298 (1), 93-96 (1992) 92183838 10 (bases 42198 to 44153) Douglas,S.E. and Reich,M.E. 1992 A dehi homolog, encoding a subunit of Mg chelatase, is located on the plastid genomes of red and cryptomonad algae J. Mar. Biotechnol. 1, 135-141 (1993) 11 (bases 82327 to 84479) Douglas,S.E. and Murphy,C.A. 1993
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Structural, transcriptional and phylogenetic analyses of the atpB gene cluster from the plastid of Cryptomonas F (Cryptophyceae) J. Phycol. 30, 329-340 (1994) 12 (bases 98901 to 114602) Wang,S.L., Liu,X.Q. and Douglas,S.E. 1994 The large ribosomal protein gene cluster of a cryptomonad plastid: gene organization, sequence and evolutionary implications Biochem. Mol. Biol. Int. 41 (5), 1035-1044 (1997) 97283757 13 (bases 61067 to 68605) Leltsch,C.E.W., Kowalik,K.V. and Douglas,S.E. 1997 The alpha gene cluster of a cryptomonad, Guillardia theta: A piece in the puzzle of chloroplast genome development J. Phycol. 1 (1998) In press 14 (bases 1 to 121524) Douglas,S.E. and Penny,S.L. 1998
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved syntenic groups confirm its common ancestry with red algae J. Mol. Evol. 48 (2), 236-244 (1999) 99128221 15 (bases 1 to 121524) Douglas,S.E. 1999
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Direct Submission Submitted (08-JAN-1998) Institute for Marine Biosciences, National Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1 Canada On Sep 15, 1998 this sequence version replaced g1:11396 g1:11297 g1:18103 g1:18281 g1:11383 g1:11407 g1:12539 g1:336730 g1:11300 g1:398949 g1:2661180. 91398949 g1:2661180. Location/Qualifiers 1..121524 /organism="Guillardia theta" /organellar="plastid:chloroplast" /db_xref="taxon:55529" /note="plastid" complement(5..86) /gene="trnY(gua)" /note="codon recognized: CAU" /product="trnA-Tyr"

gene /antlicodon(pos:50 .52,aa:Tyr)  
complement(5. .86)  
/gene="trny(gua)"  
gene /gene="trny(gua)"  
complement(181. .573)  
/gene="trp119"  
CDS complement(181. .573)  
/gene="trp119"  
/codon\_start=1  
/product="ribosomal protein L19"  
/protein\_id="AAC35594.1"  
/db\_xref="GI:3602933"  
/translaton="MEIILHLYEFVNMKSNKULIOLVESINSNKTLPOLQYGDYVRLG  
VSTLESNKEIQTEGVILSKNSGGINLTITIRAVLOGIGIERVLLSHPRKLSKTEITLL  
RKKIRRAKRYLYRTSRGKATRLQGRFN"  
729. .3188  
/gene="c1pc"  
729. .3188  
/gene="c1pc"  
/gene="c1pc"  
/product="Clp protease ATP binding subunit"  
/protein\_id="AAC35595.1"  
/db\_xref="GI:3602934"  
/translaton="MPEPTEKAIKVIIMLAOEEARLIGHNPFVTEOILILIGSGTGI  
AAVLKSMGVNLKDAVEVEKIIIGRSQFVVALEIDPFRARVLELSLEEROLGHHY  
IGEHLLILIREGSGVARVLENTALDITLVROYVITLSDTAEVATSNQTKGHHY  
TLEEPSNLTKRKAEGKLDYVPTDYGROKREIERVYIOLIGRTKNNPILIGEPVGKTAIVL  
GLAORINNDVPDILIEDKRVYITDGLILVAGTKRGEEELKTKIDELRVANNVYILV  
IDEBHYLTGAGAEAGTIDANILKPKALRAGMOCIGATTLEERYKHTEKQSALPERRO  
PVAVGSPVSEETETLLYLGRDYERKHNKLVISDELSAAAFADQYIADPFLPKAD  
LIDAEASRKLKDEENTQSPPARRELKDELEIKDKDEAVRSQDETACGLDREMEITA  
QIADINASHKSGDEENTKEVSVYTEDIDQIYAAMTGIPVNKTRSESEKILHMEETIA  
GRIGODEVAVSVSKAIRRAARKNPRIPLASFISGPTGKGTETLTKLAAYFQS  
EIAWMLDSEVSEMERHNVSKLIGSPGVVGVNEGGLTESVRRRRPVYVLPDREKCH  
PVQFNLLIOLIEDGLTDSKGTVDYFKNTLILSNVSGSLYIKPGGGGLGCLPDSBDT  
ESYOGIKRILVEDGLKGYTPRPFPLNRLDELITVLPOLIKDEYGETALMLKEVFRIRSE  
KGTQLEVARAFETHLINCGSNPIYGARPLRAVAKMLLEDTLSEEFLEAKIKEDGTAV  
DYDDGKYVYVILGKLELVAN"  
3316. .3388  
/gene="trnf(gaa)"  
/note="codon recognized: CUU"  
/product="tRNA-Phe"  
/antlicodon=(pos:3349. .3351,aa:Phe)  
3316. .3388  
/gene="trnf(gaa)"  
complement(3402. .5201)  
/gene="dnab"  
complement(3402. .5201)  
/gene="dnab"  
/codon\_start=1  
/transl\_except=(pos:complement(5199. .5201),aa:Met)  
/product="replication helicase subunit"  
/protein\_id="AAC35596.1"  
/db\_xref="GI:3602935"  
/translaton="MKVYKRFYNLKAEKHLIIQILVEKDFLOLYTKIDPVOVYFOSQ  
ILRSINILHKKKVLNLPNLSLHSEIVSEYINLTITLFDNRNPYILNIDSK  
VLDFEIRKTONSKATVSTFENPNSTVEYLLOKSNLSINNTYKNPILNIDSK  
LLELTLEIDREKTONSKATVSTFENPNSTVEYLLOKSNLSINNTYKNPILNIDSK  
VADIOSPIVYFSLSEMSKOLITLISINENINATSRREBNSISSEMLNMAITLS  
NLNITLDDENNDVLDIOSKLASLOQIYGGIQLIVDLOQLQYKESQSORHAIETSY  
TRILKILKDNNDLPVLOSLSRNERVEKRNKRPILSDLKESGCLSVATNQPYFLKSDN  
VNFSLTSLKASNHYILSATLELIPQVNRIVPYSILIKRELQDTGVKVVYELDEYIS  
VIVSFSLYVTLTGMKRIEELTVDDLVATLDQYLYNNTELODTSSVIVSFSLY  
ICMARNIVYDGTIRKNNPFVNALVHNSIEDQADVLEFLYRDSYVOHMLKQSNIDMC  
EVIYAKRHGHGICGVNLIIPNDNPFVNLIKES"  
3594. .4073  
/note="signal element intein: encodes putative intein  
spliced out of the replication helicase subunit"  
5361. .5433  
/gene="trnm(cca)"  
/note="codon recognized: GGU"  
/product="tRNA-Trp"  
/antlicodon=(pos:5394. .5396,aa:Trp)  
5361. .5433



```

EGDIFIVCEVOQLGDHCVRANISNATDGLKRGNTVEDTGVLAQVGRSTLGRJEN
VLGETIDMLGADPTSNQLPIHRSAPKFDLDTKLSIFEETIKVVDLLAPYRRGKIGL
FGAGVGKTVLIMELINNIKAKTGGVSVFEGVGGERTREGNDLYMEKESGIINELIS
ESKVALYVGMNPPGARMRGLFALTMAEYPRDVSODVLEFDNIFFRLDAGSEVS
ALLRLPSAVGYOPLASEMGLORITSTKGSITSIOAVYIPADDLDPAPATFEA
HLDATPVISRGIAKSGTYPAYDPLESTSMLOPHIVGEHYGCSOVKRTLOHRYKELQ
DIINILGIDELSPDDRLIVARAKTERLSELPFAVELFTGLPGKIVSLAKTIQGNL
ILSDLDLALSQDAFLVGDIEBASKGLFK"
gene
2345..2749
/ gene="atpe"
2345..2749
/ gene="atpe"
/ codon_start=1
/ product="ATP synthase CFI epsilon chain"
/ protein_id="CAB53103.1"
/ db_xref="GI:5748665"
/ translation="MILKILIMPDGIFMNNKAEIILPTNTGOIGILKNHPLITAL
DIGVILIRTDKRWPEIIMGGFALIKONKITLVNGAESAGTLKLVQSEARQOEAATNK
LENASKROYDALFLFKCAKARYQAQKOLVS"
2801..3424
/ gene="rps4"
2801..3424
/ gene="rps4"
/ codon_start=1
/ product="30S ribosomal protein S4"
/ protein_id="CAB53104.1"
/ db_xref="GI:5748666"
/ translation="MSRYGGRPLIIRLGLRPGFSKTIIDNRHTPPGQGMKKKASDQ
KKESQSGILKEKOKLRNGIRNGIERQILINVEARRRKSGTGEVLDLLERLDNI
IRLGFAPTTIPAAKQLNHGHINVNKNKINIPSYICKINDIISVLKNSQQLKNYLNQ
GGISLSTCLNKEKLEASINNIIPDLVLEINELLVIEYSRKL"
3603..3674
/ gene="tRNA-Asn(GUU)"
/ note="codon recognized: AAC"
/ product="tRNA-Asn"
/ anticodon="(pos:3635..3637,aa:Asn)"
3603..3674
/ gene="tRNA-Asn(GUU)"
/ gene="16S rRNA"
/ product="16S ribosomal RNA"
3807..5344
/ gene="16S rRNA"
5345..5631
/ note="Internal transcribed spacer 1, ITS1"
5432..5508
/ gene="tRNA-Ile(GAU)"
/ note="codon recognized: AUC"
/ product="tRNA-Ile"
/ anticodon="(pos:5466..5468,aa:Ile)"
5432..5508
/ gene="tRNA-Ile(GAU)"
/ gene="tRNA-Ala(UGC)"
5520..5595
/ note="codon recognized: GCA"
/ product="tRNA-Ala"
/ anticodon="(pos:5553..5555,aa:Ala)"
5320..5595
/ gene="tRNA-Ala(UGC)"
5632..8644
/ gene="23S rRNA"
/ product="23S ribosomal RNA"
5632..8644
/ gene="23S rRNA"
8645..8677
/ note="Internal transcribed spacer 2, ITS2"
8678..8796
/ gene="5S rRNA"
/ product="5S ribosomal RNA"
8678..8796
/ gene="5S rRNA"
complement(9188..10267)
/ gene="mind"

```

```

CDS
complement(9188..10267)
/ gene="mind"
/ codon_start=1
/ product="cell division inhibitor mind"
/ protein_id="CAB53105.1"
/ db_xref="GI:5748667"
/ translation="MNKLHYFINNI FNLIVYLYSLYFKEDKIKRRLSNMTKKQBNY
KEOLIKKPEPERKIKKEDLEQLQKPSSEXYNTLEDLDGSDLEPRPVITVSGK
GGVKKTTTANIGSMISARFGRVALIDADIGLRNLDLLENTRTTFAMDIIEGRCR
LDOLAVREKRWKNLALAVASKHGYANTQOHLMOQLVYSIKETLINSITLIDCPAGIDV
GFNAIAPAGEAIIYVTPETRIADPADPAGVGLLEANTIVDPKLLIRRMPIQSTNM
LSINDVETLIGLPLGAIPEEDTNVITISNKEPILVLDKLTLSGIAFBNARRLIGKE
DYFVDLIDTPKSIILIKTKQKFPWGE"
complement(10361..10434)
/ gene="tRNA-Arg(ACG)"
/ note="codon recognized: CGU"
/ product="tRNA-Arg"
/ anticodon="(pos:10398..10400,aa:Arg)"
complement(10361..10434)
/ gene="tRNA-Arg(ACG)"
10685..10828
/ gene="rpl32"
10685..10828
/ gene="rpl32"
/ codon_start=1
/ product="50S ribosomal protein L32"
/ protein_id="CAB53106.1"
/ db_xref="GI:5748668"
/ translation="MAVPKRTSKSKNLRKSQKQAQFVQAKKALAKAKTVLKLLN
KDN"
10950..11717
/ gene="cysT"
10950..11717
/ gene="cysT"
/ codon_start=1
/ product="sulfate transport permease protein"
/ protein_id="CAB53107.1"
/ db_xref="GI:5748669"
/ translation="MLGNHLEFIFVLPLEALFSLIKNSMKDILEKAVDPAICAYGF
TIKMLIAALFNSIEFGLITVIRYERFGKKFIDAADVDPALPTSAVAGTIAVYG
NOGVAGREIKMGNIQIITYKGVILAMIEVSPFYIRSLQVYLOGIDGLGEANACG
ASSRQTRIVPTPLVPALVGTFLTSRAIGFQSVVMISSNLPDLVLSVLIYOS
LEQYDFGASVIGAVILMILILIFLINTQAQFSRR"
11892..13652
/ gene="ycfI"
11892..13652
/ gene="ycfI"
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="CAB53108.1"
/ db_xref="GI:5748670"
/ translation="MSLSTHIRDYVEVLTGVTEASGNPLQAKLTSESLLYVLKCSQ
EVLQIISFQWRNPSLPIKIPAIYESITGOTPEBALDPEVHLGONPVYIAGLNS
APFALPESAIFHVSIRRLTQGVPAATYFSGYIIGQILFASCVYEGODIILPMTL
EPLNYINGLILSRIITISMRESIALEETMHPKRYKTYIRFLIACWEGOSIFQFLS
NITPSANPTIIOGFANGLNGLVONSTYIGGLIGASAPFLFWMLPLKLTQTYLVH
TLVYHQIVATVNOICFLSALTLSFALPLPYAAYVYLVPGLGFPEDNALSTVFTHS

```

Query Match 22.0%; Score 214.8; DB 33; Length 22085;  
 Best Local Similarity 57.8%; Pred. No. 9.3e-51;  
 Matches 428; Conservative 0; Mismatches 297; Indels 15; Gaps 2;

```

QY 172 ccgcglatcgcgtatatacctcgcgaaaggcgtgtggaagaagacacaccgcga 231
DB 9991 CCTAGGGTTATAGTAATTAACGCTGTGAAAGTGTAGTAAACGACACTCTGCT 9932
QY 232 aatgcgctcctcctcgcgcgtatcctcgaagtgctgcacatgcgcgaacct 291
DB 9931 AATTGGGAATGTCATATGCAAGATTGGATATCTGTAGCATTTATATGATCTGATATA 9872
QY 292 ggtccgtaaacctcgatcctcctcctagggttagagaatcgagtcacatcacatcgctc 351
DB 9871 GGGTTACGTAAATTATAGATTATAGCATTAAGCAATGCAATTAAGCTTTTACACCAATG 9812

```

OY	352	gaaggttaaaacggagaaatctgtcgtctccgataccaagctctggtaacgtttgacg	411
Db	9811	CAATATTATTGAAGGTCGTTGTCGACTAGATCAAGCGTTGGTACCTGAAGAAGCGTTGGAAA	9752
OY	412	aatttcgaatctgcatlgtatcctaaactbagaicgaacatccgatbgyggaatttgtyggt	471
Db	9751	AATTTAGCATTTATWGCGACGTTTCAAA-----AAACACCCAAAATAATATGTAC	9702
OY	472	aaagcattggaattgctctgtggaatcgtttgaaactagaccggaagaagttaccggaatttc	531
Db	9701	GCACACACATATGCGCAATTAGTTTTCATTAATAAAGAACTAGGTAT-----CAATTCA	9647
OY	532	atcacatcagatgtccctccaggaatccggaatcccggaattcaaccgccatctaccgcggc	591
Db	9646	ATATTAAATTGATTGGCCCGCAGGTATGTGATGTTGTTTAAATGCAATAGCACCCAGCA	9587
OY	592	aatgaagcagctctctgttaacaactcccggaatatacacgcgtttaaggaaatgtcatlaaggtc	651
Db	9586	CAAGAAGCATTTATTTGTACAAACCTCGAATAATACAGCTTTTCGTGATGCGAATAGAGTG	9527
OY	652	acgggctctgtcagaatbcsatcggaatccagagataataagaatgattgtgaaacgaagtga	711
Db	9526	GCTGATTTATTTGAAGCGATACAAATTTGTGATACAAAAATTAATTAATTCGCTGTCTG	9467
OY	712	actgatatgatataaaggaaagatagatagctgcgtgtgtgagtgtgcaggagatgttggga	771
Db	9466	ATGGATATTCATTCCAANAATGTCACAAATGTTATTCATATATGATCTTCCAAAGAACATTTAGGT	9407
OY	772	tgtcatatgctctgtgttaactctctgaagaatccgaagttatctcgaagcaagaatccgaagc	831
Db	9406	ATTCGATTTATTTGGCCCATCTACTCGTAAGATACAAATGTATTTATTTCCACTAATAAAGGA	9347
OY	832	ttccgcctgtctcgaataagcctccactacgcttcggaagttgagtcgtttgacgaagcgctc	891
Db	9346	GAACCTTTAGTCTTGATGATAAAAAATTAACGCTTTCAGGTATTCATTTGAAGAACGCCGCG	9287
OY	892	tggagacatcglttgagcaaga	911
Db	9286	CGTAGATTAATATGCAAAAGA	9267

LOCUS	SYNCLG	135638 bp	DNA	BC1	13-FEB-1999
DEFINITION	Synechocystis sp. PCC6803 complete genome, 24/27, 3002966-3138603.				
ACCESSION	D64005 AB001339				
VERSION	D64005.1 GI:1001779				
KEYWORDS	UmuC; ABC transporter; DNA polymerase I; P-methylase; Pled; UmuC protein; beta-lactamase; bromoperoxidase; cadmium-transporting ATPase; cation efflux system protein; cation-transporting ATPase PacI; cytochrome b6-f complex subunit PetM; eukaryotic protein kinase; glutamate dehydrogenase (NADP+); heterocyst to vegetative cell connection protein; magnesium and cobalt transport protein; phosphoenolpyruvate synthase; phosphofructokinase; polyphosphate kinase; pre-B cell enhancing factor; protein kinase; pyridoxal phosphate biosynthetic protein PdxA; pyruvate oxidoreductase; regulatory components of sensory transduction system; salt-stress induced hydrophobic peptide; sensory transduction histidine kinase; septum site-determining protein Minc; septum site-determining protein Minc; septum site-determining protein Minc; transposase; repressor SmcB; transfer RNA-guanine transglycosylase; transposase Yc135.				
SOURCE	Synechocystis sp. (strain:PCC6803) DNA.				
ORGANISM	Synechocystis sp.				
REFERENCE	Bacteria: Cyanobacteria; Chroococcales; Synechocystis.				
AUTHORS	1 (sites) Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugita, M., and Tabata, S.				

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS
Sequence analysis of the genome of the unicellular cyanobacterium <i>Synechocystis</i> sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome	DNA Res.	2 (4), 153-166	(1995)	2 (bases 1 to 135638)
Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.	JOURNAL	97061201		Location/Qualifiers
				1..135638

[illegible]











BSMINCD 1900 bp DNA BCT 02-MAR-1993  
 LOCUS BSMINC B.subtilis minC, mind, and spoIVFA (partial) genes.  
 DEFINITION B.subtilis minC, mind, and spoIVFA (partial) genes.  
 ACCESSION 215113  
 VERSION 215113.1 GI:49307  
 KEYWORDS minC gene; mind gene; spoIVFA gene; sporulation operon.  
 SOURCE Bacillus subtilis.  
 ORGANISM Bacillus subtilis.  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.  
 1 (bases 1704 to 1900)  
 Cutting, S., Roels, S. and Losick, R.  
 Sporulation operon spoI and the characterization of mutations  
 that uncouple mother cell from forespore gene expression in  
 Bacillus subtilis  
 J. Mol. Biol. 221 (4), 1237-1256 (1991)  
 JOURNAL MEDLINE 92046062  
 2 (bases 1 to 1900)  
 Lee, S. and Price, C.W.  
 The minC locus of Bacillus subtilis lacks the mind determinant  
 that provides topological specificity to cell division  
 J. Microbiol. 7 (4), 601-610 (1993)  
 JOURNAL MEDLINE 93211302  
 3 (bases 1 to 1900)  
 Price, C.W.  
 Direct Submission  
 Submitted (15-SEP-1992) Chester W. Price, Department of Food  
 Science, University of California, Davis, Department of Food  
 Science, University of California, Davis, California, 95616, USA  
 Location/Qualifiers  
 1. 1900  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /db\_xref="taxon:1423"  
 /clone="PSY, PSY4, PSY13"  
 173. 180  
 /note="putative sigma-H recognition sequence"  
 194. 198  
 /note="putative sigma-H recognition sequence"  
 212. 218  
 /gene="minC"  
 join(212..218, 224..904)  
 /gene="minC"  
 224..904  
 /gene="minC"  
 /function="Cell division inhibitor"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MinC"  
 /protein\_id="CAA78817.1"  
 /db\_xref="GI:580892"  
 /translation="MKTKKOQVYTIKGTNGLTLHLDDACSPPELLDGLNMLSIEQY  
 TGGKGIKIVHYKLNRLPKREBEOITELIAKKDLFVSISEVITTKKKAQRIRE  
 AAVISVKIPIVSGQVQLVKGDLILIGVNGGIVRGGNIFVLSGLKGIHAGNGNN  
 QAVIASSEMPLQOLRINHLNRPDIHOKNEMECAYLDTGNNVIERLQHLALRPD  
 LRLREGGM"  
 893..897  
 /gene="mind"  
 893..897  
 /gene="mind"  
 906..1712  
 /gene="mind"  
 906..1712  
 /gene="mind"  
 /function="ATPase activator of MinC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Mind"  
 /protein\_id="CAA78818.1"  
 /db\_xref="GI:580893"  
 /translation="MGBAIVITSGKGVGKTTTSANLGTALAIIGKRVCLVPTDIGLR  
 CDS  
 RBS  
 gene  
 gene  
 gene  
 CDS

LDVVMGLNRIIYLDVVEGRCKMHAQVKKRFDLLYLPMAQTSDBKTAPEQ  
 IKNNVBLEKOEPIVITDCPAIGOGKKNAGADKAIVTTPEISAVDADRIIGL  
 EORENIEPRLVIRIIRNHLNLMKNGDMDIDEIVOHLSIDLGIYADDEVIKASHGE  
 PIAMPKNNRASIATRNIRILGSEVPLQVLEEDKNGKMAIKSPFGVRS"  
 terminator  
 1737..1769  
 1862..1900  
 /gene="spoIVFA"  
 1862..1900  
 /gene="spoIVFA"  
 /function="coupling gene expression"  
 /citation=11  
 /codon\_start=1  
 /transl\_table=11  
 /product="SpoIVFA"  
 /protein\_id="CAA78819.1"  
 /db\_xref="GI:49310"  
 /db\_xref="SWISS-PROT:P26936"  
 /translation="MSHRADIRKLE"  
 BASE COUNT 600 a 341 c 445 g 514 t  
 ORIGIN  
 Query Match 13.2%; Score 128.8; DB 2; Length 1900;  
 Best Local Similarity 53.0%; Pred. No. 4,3e-26;  
 Matches 355; Conservative 0; Mismatches 297; Indels 18; Gaps 3;  
 QY 182 tcgtatcacctccggaagcggtgtgtgaagacgacacccagcaatgtcgtc 241  
 DB 919 TCGTAAATTAACCTCGGGAAGCCGAGTAGTACACACACATCTGCAACCTCGGTA 978  
 QY 242 tctctcgcgtcgttcggttctcagttgtcgcacgtgagccgacctgtctcgt 301  
 DB 979 CCGCCTTAGCCATTAGGGAAGCCGATGCTTAGTATGATCTGATATAGAGACTCGCA 1038  
 QY 302 acctcgatcctcctctaggttagaagatcgatcactcgtcgtcgaagtataa 361  
 DB 1039 ACCTGATGCTTGAATGAGGCTTTGAATAATGATTTTACATCTCGTAGAGCTTGTAG 1098  
 QY 362 acgagatgtcgtcgtcgtcgaagctcgtcgtatagaagcgtgtgtcgaattcgaat 421  
 DB 1099 AGGCGAATGCAAAATGATGACGCGCTCGTAAAGACAAACCTTTGATGATCTGCTCT 1158  
 QY 422 tgcattatatacctaactagatcgaaacttcggtggttgggtgaagcattgg 481  
 DB 1159 ATTTA---ATGCCCGCTGCTCAACGACGATTAAGACAGCTGTGCTCTGAAACAATTGA 1215  
 QY 482 aatgctgtgtgattgattgtgaagaactagacccgaaggttcaacggattcatcatc 541  
 DB 1216 AAAACATGCTCCAGAGCTCA-----ACAGGAAATTTGACTATGTCATCATAG 1263  
 QY 542 attgtcctcgaagatcgatcgcggaatcataacccgacttactccggcggaatgaagcag 601  
 DB 1264 ACTGTCCTCCCGCAATCGACGAAGGGGTACAAAATCCCGTTCCGGAGCTGATTAAGCGA 1323  
 QY 602 tctcgttaacaactccggaatataacagcgtttaaggatcgttgaaggattgttgc 661  
 DB 1324 TTGTGTCTACTACGCTGGAATCTTCAGCTGTTCTGATGCTGACCGTATTTATGAGACTGC 1383  
 QY 662 tagaatcgatcgatcagagaga---tataaagaatgtgtgaacagagtgagaactgata 718  
 DB 1384 TGGAGCAAGAGGAAATATTGTAACCGCCGCGCTGCTTTAACAAGAAATCAGAAATCACC 1443  
 QY 719 tgattaaagagagatagatagatcagttgttagatgtcgaagagatgttggtatgtcat 778  
 DB 1444 TGATGAAAAAAGCGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1503  
 QY 779 tgcctgtgtaattcctgaagatcctgaaggttatacgaagcaagaatcgaaggttcgcgc 838  
 DB 1504 TCGTCGGAATCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1563  
 QY 839 tctgtctgaa 848  
 DB 1564 TTGCGATGGA 1573

```
RESULT 15
BACREM1N 4389 bp DNA BCT 26-APR-1993
LOCUS BACREM1N subfills (mreb) gene complete cds, (mrec) gene complete
DEFINITION cds, (mred) gene complete cds, (mnc) gene complete cds, (mind)
gene complete cds.
ACCESSION M95582
VERSION M95582.1 GI:143211
KEYWORDS mnc gene; mind gene; mreb gene; mrec gene; mred gene; shape
SOURCE determining protein.
ORGANISM Bacillus subtilis (strain W168) DNA.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE Varley,A.W and Stewart,G.C.
AUTHORS The Diviv Region of the Bacillus subtilis Chromosome Encodes
TITLE Homologs of Escherichia coli Septum Placement (MinCD) and Cell
JOURNAL Shape Determinants
MEDLINE J Bacteriol. 174, 6729-6742 (1992)
FEATURES
source location/Qualifiers
1. .4389
/organism="Bacillus subtilis"
/strain="W168"
/db_xref="taxon:1423"
/map="247 degrees"
306. .1339
/gene="mreb"
306. .315
/gene="mrec"
326. .1339
/feature="putative"
326. .1339
/feature="putative"
/codon_start=1
/transl_table=11
/function="cell shape determination"
/protein_id="AA22605.1"
/db_xref="GI:143212"
/transl_table="MGI:GARDIGDIDLGANTLVFKGKGIIVREPSVALOTDTKSI
VAVGNDKAMIGRPNVVALRPMKGDVADYETATPMKXYINOAIRKKGMAKRPY
VMVCYPSITVAEERAVDADTRQACARDAIPIEPPAAITGALPWEPTGSAVNDIG
GGTTEVAITISGGIVTSQIRVAGDEMDALINIRKTYNLMIIGDTAEIKMETISA
EAPESDKMEIRGRDLTGLPTIETITGCKEMALADVTIVEAVKSTLEKTPPLA
ADIMRGIIVLGGGALLRNLDKVISSEETKMPVLIADPDCAVIGTGAKLEHILFKG
KTR"
1355. .2242
/gene="mrec"
1355. .1362
/gene="mrec"
1370. .2242
/gene="mrec"
/feature="putative"
/feature="putative"
/codon_start=1
/transl_table=11
/function="cell shape determining"
/protein_id="AA22606.1"
/db_xref="GI:143213"
/transl_table="MNRKRLMLLLCIILVAMIGFSLKGNNTTWPEKVICPTGVF
ONITTPAEFAGIENINDLKNTYKENEKRLKEDGOTQYEAKELOLEENSLDE
LGHVSKIDYPIATVARSPPDMAKQYITIKGQOONAKMAVATNEGALGKIKS
SGLNNTSAVOLLSPTDRNNRYATKISGKSGKYGILGEGYKREKRLKMTIERDK
ODVKGDILIESGCTGVPPEGLTIGEVVDIESDSIGLTRVATYKPADLTLNNVIV
NRDVPYDVEEGS"
2224. .2231
/gene="mred"
2224. .2757
/gene="mred"
2239. .2757
/gene="mred"
CDS
```

```
/note="putative"
/codon_start=1
/transl_table=11
/function="cell shape determining"
/protein_id="AA22607.1"
/db_xref="GI:143214"
/transl_table="MKRFLDPVWMLVFSASISITFDLVHPFPYTDVQVLAIRPLMLVL
IFMSAFINQKHAMIVGFTFCGLYDMNTSLIGVYFPGFAGLCYLASKAFKVLHTNAFV
VILAVLAVCLLEFFVFIQSLIHKDINTFNGFVLDREPIILLINAAALILVPEFL
FMSLKELEDE"
2798. .4298
/gene="mind"
2798. .2803
/gene="mind"
2810. .3490
/gene="mind"
2810. .3490
/gene="mind"
/feature="putative"
/feature="putative"
/codon_start=1
/transl_table=11
/function="septum site determination"
/protein_id="AA22608.1"
/db_xref="GI:143215"
/transl_table="MKTQKQYVTITKTRKGLTLHLDDACSFDELLDQNLSTEOY
TDGKQKISVAVKLGKRNPLKKEOEOLLELASKKDLVHSDISEVITKKEOQIRE
AEITSVKIYRSGOVLOYKGLDILLIGDVPNGGTAVAGNITVIGSLKGTARHFGN
OAVTASSEMPLPTOLRINIVNLNRPDHIQKGNMECAVILDTGNNVIERLOHLHLRPD
LRLLEGH"
3477. .3483
/gene="mind"
3492. .4298
/gene="mind"
/feature="putative"
/feature="putative"
/codon_start=1
/transl_table=11
/function="septum site determining"
/protein_id="AA22609.1"
/db_xref="GI:143216"
/transl_table="MGEAIVITSGKGVGKTTTSANLGTALAILGRVCLVDTIGLR
NLDVVMELENIIDVLDVVEGRCKMHQALVKDKFDDLLVMPAAGSDKAVAPED
IKNWOELKOEFDVILIDCPAGIEGYNNAVAGADKATVMTPEISAVRADRIIGL
EOENITPEPRIVVVRIRHMLKNGDPTMDIDELIYHLSLIDIGTIVADDEVIKASHNGE
PIANDPKNRASIAIRNARILRLLGESVPLQVLEQKGMMAKIKSFVGRS"
4323. .4361
/standard_name="transcription terminator"
/feature="putative"
BASE COUNT 1343 a 812 c 1056 g 1178 t
ORIGIN
Query Match 13.2%; Score 128.8; DB 2; Length 4389;
Best Local Similarity 53.0%; Pred. NO. 4.8e-26;
Matches 355; Conservative 0; Mismatches 297; Indels 18; Gaps 3;
terminator
182 tcgtatcaccctcggaaagcggtgttggaagaagcaaccacccggaataatgctgcgc 241
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3505 TCGTATACCTTCGGGAAAGCGGAGTAGAGACACACATCTGCGAACCTGGGTA 3564
242 tctctcgcctcgtacggttctcagtcgcatcagcgcgcgaccttggtctcgcga 301
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3565 CCGCTTAGCCATTTAGGGAAGCCGATGCTTAGTAGATACGTATATAGCAGCTCGCA 3624
302 acctcgatctcctctaagggtaaggaactcgaataacacctgcgcgcgagtataa 361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3625 ACCTTGATGTTGTAATGGCTCTTGAAATAGATTAATTACATCTGTTAGACGTTGTAG 3684
362 acggagatcgtcgtcgcgaagctcgtcgaataagcgttcgtaacattcgaat 421
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3685 AGGGGATGCGAANAATGCATCAGCGCGCTGTAAGACAAAGCTTTCGATCTCTCT 3744
422 tgcctatatactcaaacctcgatcgaacacctcgaatggaatttggtgaagcattg 481
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db	3745	ATTTA---ATGCGCCGCTGTCACAAACAGCGATTAAGACAGCTGTTGCTCCGTAACAAATT	3801
OY	482	aatgcctctgtgatctgccttgtaaaactatagaccgaagttcaaccgattcaatcaatcgc	541
Db	3802	AAAACATGGTCCAAAGGCTCA-----AACAGCAATTGGACTATGCTATCATAG	3849
OY	542	attgtccctcaggaatcagatgcgcggaattcaaccgcctattccgcggaatgaacag	601
Db	3850	ACTGTCCTCCCGGAATCGACAAAGGTTACAAAATGCGTTCCGGAGCTGATTAACGA	3909
OY	602	tctctgtaacaactccgcgatatacagcgttlaagggatctgtaggtatcaggtttgt	661
Db	3910	TGTGGTCACTACGGCTGAATCTCAGCTGTTGCTGATGCTGACCGGATTAATAGAGATGC	3969
OY	662	tagaatgcgatygaatcagaga---tataaagatgatttgaacagaytgaactgata	718
Db	3970	TGGACCAAGAGAAATATTGAACGCCGCCGCTGCTTTTGAACAGATAGGAATACC	4029
OY	719	tgatctaaagagagatatgatgtcagtggttagatgtgcagagatgttgagttgcat	778
Db	4030	TCATGAAAACGCTGCACAGATGATATGCACAAATCGTACAGCATCTTCGATCGATT	4089
OY	779	tgccttggttlaatccctgaagaattctgaagttatcgaagcaagaaatcgaaggtttccg	838
Db	4090	TGCTCGGAATCGTGCTGATGATGATGAAGTCAATTAAAGCTTCCAATCATCGGAAACGA	4149
OY	839	tgtcttcgaa 848	
Db	4150	TTGCGATGGA 4159	

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2001, 03:42:11 ; Search time 118.73 Seconds

(Without alignments)  
3094.400 Million cell updates/sec

Title: US-09-553-431-1

Sequence: 1 atggcgtctctgagatgtt.....tcttcttcttcttgcgcgc 978

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq.36:\*

- 1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.8	4.4	885	19	X30457
2	42.8	4.4	891	18	H. pylori cellular
3	42.4	4.3	513	18	H. pylori cytoplasm
4	42.4	4.3	6761	20	H. pylori cytoplasm
5	42.2	4.3	1236	18	Polymerase chain reaction
6	42.2	4.3	1275	18	H. pylori cytoplasm
7	39.2	4.0	780	19	H. pylori cytoplasm
8	39.2	4.0	1107	19	Nucleotide sequence
9	38.4	3.9	1664976	19	Methanococcus jann
10	38.2	3.9	1012	20	Nucleotide sequence
11	37.4	3.8	30600	20	Human MTH1 relate
12	35.6	3.6	6659	21	DNA encoding a GON

13	35	3.6	1590	20	X07125
14	35	3.6	7778	18	H74359
15	34.6	3.5	534720	19	V30458
16	34.6	3.5	536165	19	V30459
17	33.6	3.4	44377	18	T78508
18	33.6	3.4	44377	18	T80414
19	33.4	3.4	1610	20	X13498
20	33.4	3.4	58407	19	V21210
21	32.8	3.4	462	21	V42709
22	32.8	3.4	1128	18	V31211
23	32.8	3.4	10401	18	V74460
24	32.8	3.4	12022	20	X13054
25	32.2	3.3	1666	21	Z86795
26	32.2	3.3	26281	21	Z60929
27	32	3.3	15062	19	V44232
28	32	3.3	116277	20	X20249
29	31.8	3.3	400	18	V77264
30	31.8	3.3	7032	20	V99801
31	31.8	3.3	7494	21	Z56579
32	31.8	3.3	7494	18	V74380
33	31.8	3.3	9223	19	V52192
34	31.8	3.3	1664976	19	V21209
35	31.4	3.2	1109	18	V74533
36	31.4	3.2	72928	20	Z18355
37	31.2	3.2	1967	20	V64414
38	31.2	3.2	26016	19	V57272
39	31	3.2	1690	18	V74797
40	30.8	3.2	9897	19	V52143
41	30.8	3.1	1098	18	T67896
42	30.8	3.1	3766	19	V69424
43	30.8	3.1	3766	21	Z89989
44	30.8	3.1	7664	20	X13312
45	30.6	3.1	5678	20	X13570

#### ALIGNMENTS

RESULT 1

ID X30457 standard; DNA: 885 BP.

AC X30457;

DT 08-JUN-1999 (first entry)

DE H. pylori cellular protein ORF 06p11202\_4569693\_c2\_28.

XX Vaccine probe; diagnostic; ORF; cell envelope protein;

KW secreted protein; cellular protein; ds.

XX Helicobacter pylori.

OS Helicobacter pylori.

XX WO9818323-A1.

PN 07-MAY-1998.

PD 28-OCT-1997; 97WO-US19575.

PF 14-JUL-1997; 97US-0891928.

PR 28-OCT-1996; 96US-0739150.

PR 06-DEC-1996; 96US-0759739.

XX (ASTR) ASTRA AB.

PA Alm RA, Smith D;

PI WPI, 1998-271811/24.

DR P-PSDB; Y10990.

DR Helicobacter pylori nucleic acids and proteins - used to develop

PT products for the detection, prevention and treatment of H. pylori

PT infections

XX

PN WO9640893-A1

PN WO9640893-A1



XX (ASTR ) ASTRA AB.  
PA Berglindh OT, Smith D, Mellgaerd BL.  
XX WPI: 1997-052306/05.  
XX P-PSDB; W20443.  
XX  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX  
PS Claim 9; Pages 267; 1481pp; English.  
XX  
CC The present sequence encodes a Helicobacter pylori cytoplasmic protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds, useful  
CC as potential H. pylori life cycle activators or inhibitors. The genomic  
CC sequence of H. pylori (ATCC 55679) was determined from overlapping  
CC contigs generated by mechanically shearing the bacterial DNA. The  
CC sequences were analysed for ORF of at least 180 nucleotides, and the  
CC predicted coding regions defined by computer evaluation. To identify  
CC likely H. pylori antigens for vaccine development, the amino acid  
CC sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
XX  
SQ Sequence 1236 BP; 398 A; 239 C; 282 G; 317 T; 0 other;  
XX  
Query Match 4.3%; Score 42.2; DB 18; Length 1236;  
Best Local Similarity 52.6%; Pred. No. 0.002;  
Matches 92; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
XX  
QY 155 aactgcgcgagaaacgcgcatcgtatcgtatcacctccggaagcggtgtgaa 214  
DB 404 atctgtctaaacatacgaatgtgtcatgataagcagcggtgtggtgta 463  
QY 215 agacgacacacccgcaaatgtcgtctctctcgtcgttcgttcagttgtcg 274  
DB 464 aaagcacaccacgcgtgaatttaagcatcgtcttagcgaatttaacaaagctgggc 523  
QY 275 ccattgagccgacgtgtctcgttaacctgacctctcctcctaggttagaa 329  
DB 524 tactagacgctgattgtatgtgccttaatacctcctagatgaggtcttgcaaa 578  
XX  
RESULT 6  
T68266  
ID T68266 standard; DNA; 1275 BP.  
XX  
AC T68266;  
XX  
DT 22-JUL-1997 (first entry)  
XX  
DE H. pylori cytoplasmic protein ORF hpse1521lorf13.  
XX  
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.  
XX  
OS Helicobacter pylori.  
XX  
FH Key Location/Qualifiers  
FT 1..1275  
FT CDS  
FT /\*tag= a  
FT /note= "no stop codon given in sequence"  
XX  
PN M09640893-A1.  
XX  
PD 19-DEC-1996.

XX  
PF 06-JUN-1996; 96WO-US09122.  
XX  
PR 01-APR-1996; 96US-0630405.  
PR 07-JUN-1995; 95US-0487032.  
XX  
PA (ASTR ) ASTRA AB.  
XX  
PI Berglindh OT, Smith D, Mellgaerd BL.  
XX  
XX WPI: 1997-052306/05.  
XX P-PSDB; W21013.  
XX  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX  
PS Claim 9; Pages 1001; 1481pp; English.  
XX  
CC The present sequence encodes a Helicobacter pylori cytoplasmic protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds, useful  
CC as potential H. pylori life cycle activators or inhibitors. The genomic  
CC sequence of H. pylori (ATCC 55679) was determined from overlapping  
CC contigs generated by mechanically shearing the bacterial DNA. The  
CC sequences were analysed for ORF of at least 180 nucleotides, and the  
CC predicted coding regions defined by computer evaluation. To identify  
CC likely H. pylori antigens for vaccine development, the amino acid  
CC sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
XX  
SQ Sequence 1275 BP; 410 A; 246 C; 289 G; 330 T; 0 other;  
XX  
Query Match 4.3%; Score 42.2; DB 18; Length 1275;  
Best Local Similarity 52.6%; Pred. No. 0.002;  
Matches 92; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
XX  
QY 155 aactgcgcgagaaacgcgcatcgtatcgtatcacctccggaagcggtgtgaa 214  
DB 443 atctgtctaaacatacgaatgtgtcatgataagcagcggtgtggtgta 502  
QY 215 agacgacacacccgcaaatgtcgtctctcgtcgttcgttcagttgtcg 274  
DB 503 aaagcacaccacgcgtgaatttaagcatcgtcttagcgaatttaacaaagctgggc 562  
QY 275 ccattgagccgacgtgtctcgttaacctgacctctcctcctaggttagaa 329  
DB 563 tactagacgctgattgtatgtgccttaatacctcctagatgaggtcttgcaaa 617  
XX  
RESULT 7  
V90595/C  
ID V90595 standard; DNA; 780 BP.  
XX  
AC V90595;  
XX  
DT 18-FEB-1999 (first entry)  
XX  
DE Nucleotide sequence of clone Y108-1.ASM from cluster 32.  
XX  
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;  
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.  
XX  
OS Helicobacter pylori.  
XX  
PN M09849314-A2.  
XX  
PD 05-NOV-1998.

PE 27-APR-1998; 98WO-US008487.  
XX  
XX 14-OCT-1997; 97US-0061958.  
PR 25-APR-1997; 97US-0045107.  
XX  
XX  
PA (GENE-) GENELABS TECHNOLOGIES INC  
XX

PI Chow TP, Fry KE, Lim MY, Mcatee CP,  
XX  
DR WPI; 1999-009433/01.  
XX

**Pt** New Helicobacter pylori antigens and related nucleic acid sequences  
**Pt** - useful in serological diagnosis and protective vaccines, providing  
**pt** long-lasting immune response  
**xx**  
**xx**

**Claim 27:** Page 125; 402pp; English.  
**PS**  
**XX**

The present sequence encodes a Helicobacter pylori antigenic protein that is characterised by immunoreactivity with H. pylori-positive antisera. The proteins are highly immunogenic and induce a long-lasting immune response that persists even after antimicrobial treatment. In antibody-detection assays, on sera, plasma, urine, saliva etc., they are highly sensitive and specific. The specification also describes 69 previously unrecognised immunogenic cluster families. H. pylori antigens are used to detect H. pylori-specific antibodies, for diagnosing infection or to confirm eradication of infection, and in vaccines to protect against H. pylori infection and related diseases (gastritis, peptic ulcer, gastric adenocarcinoma/lymphoma).

Sequence 780 BP; 218 A; 157 C; 142 G; 263 T; 0 other;

Query Match	4.08;	Score 39.2;	DB 20;	Length 780;
Best Local Similarity	53.28;	Pred. No. 0.014;		
Matches	83;	Conservative	0;	Mismatches 73;
			Indels	0;
			Gaps	0;

QY	174	gcgtatcgcgttcaacccttcggaaaaagcgcgcgttgcggaaagcagacacccaccgaaa	23
Db	296	GCATGTACTGATGATGAAGCTAGGTAAGCGCGCTGTGGTAAAGCACCCACACGGTAA	23
QY	234	tcctcgatcctcctcctcgtctcgtcttcctcgttctcgttgcgcattgaagcgcgaacttgg	29
Db	236	TTTAAAGCATGCCCTTTAACGATTAAACCAAAAAGTGGCGGTACTAGCGCGCTGATGTGTA	17
QY	294	tcctcctaacctgcgatactcctccctcaggatgttagaaga	329
Db	176	TGGCGCTAATATCCCTGAATGATGGCGCTTACCAAAA	141

PA	XX	OS	Trilicium durum.
PR	XX	PN	FR2757538-A1.
PR	XX	PD	26-JUN-1998.
PR	XX	PF	18-DEC-1997; 97FR-0016059.
PR	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA	XX	PI	Glutenin gene: durum wheat; low-molecular-weight,
PA	XX	PI	transgenic durum wheat; ss.
PA	XX	PI	Trilicium durum.
PA	XX	PI	FR2757538-A1.
PA	XX	PI	26-JUN-1998.
PA	XX	PI	18-DEC-1997; 97FR-0016059.
PA	XX	PI	19-DEC-1996; 96IT-MI02663.
PA	XX	PI	(ITU-V.) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PA	XX	PI	09-OCT-1998 (first entry)
PA	XX	PI	DNA encoding a durum wheat glutenin protein.
PA			

XX Cardelli LE, D'Ovidio R, Marchitelli C, Porceddu E  
PI  
XX  
DR WPI, 1998-365055/32.  
DR P-PSDB; W62647.  
XX

**Claim 1:** Page 13; 18pp; French.

CC The present sequence represents a glutenin gene, and is isolated from  
CC the genomic DNA of Triticum durum L. The gene codes for a  
CC low-molecular-weight glutenin protein and can be used to produce  
CC transgenic durum wheat plants with "better quality characteristics"  
CC (no details given).

sequence 1107 BP; 368 A; 377 C; 152 G; 210 T; 0 other,

Query Match	4.0%;	Score 39;	DB 19;	Length 1107;
Best Local Similarity	44.8%;	Pred. NO. 0.019;		
Matches 150;	Conservative 0;	Mismatches 185;	Indels 0;	Gaps 0

[illegible]

RESULT	9
V21209/c	
ID	V21209 standard; DNA; 1664976 BP.
XX	
XX	V21209;
AC	
XX	
DT	10-NOV-1998 (first entry)
XX	
DE	
XX	Methanococcus jannaschii circular chromosome.
XX	
KW	Methanococcus jannaschii; methanogenic archaeon; circular chromosome
KW	genome; autotrophic; extrachromosomal element; identification; ds.
XX	
OS	Methanococcus jannaschii.
XX	
PN	W09807830-A2.
XX	
PD	26-FEB-1998.
XX	
PF	22-AUG-1997; 97WO-US14900.
XX	
PR	22-AUG-1996; 96US-0024428.
XX	
PA	(GENO-) INST GENOMIC RES.

[illegible]

XX 14-Oct-1997; 97US-0061958.  
PR 25-APR-1997; 97US-0045107.  
XX  
XX (GENE-) GENELABS TECHNOLOGIES INC.  
PA  
XX  
XX Chow TP, Fry KE, Lim MY, McAttee CP;  
PI  
XX  
XX WPI: 1999-009433/01.  
DR  
XX  
XX New Helicobacter pylori antigens and related nucleic acid sequences  
PT - useful in serological diagnosis and protective vaccines, providing  
PI long-lasting immune response  
XX  
XX  
PS Claim 23; Page 297; 402pp; English.  
XX  
XX The present sequence encodes a Helicobacter pylori antigenic protein  
CC that is characterised by immunoreactivity with H. pylori-positive  
CC antiserum. The proteins are highly immunogenic and induce a long-lasting  
CC immune response that persists even after antimicrobial treatment. In  
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are  
CC highly sensitive and specific. The specification also describes 69  
CC previously unrecognised immunogenic cluster families. H. pylori antigens  
CC are used to detect H. pylori-specific antibodies, for diagnosing  
CC infection or to confirm eradication of infection, and in vaccines to  
CC protect against H. pylori infection and related diseases (gastritis,  
XX peptic ulcer, gastric adenocarcinoma/lymphoma).  
XX  
SQ Sequence 1012 BP; 293 A; 188 C; 197 G; 333 T; 1 other;

	Query Match	3.9%	Score 38.2	DB 20	Length 1012	
	Best Local Similarity	52.9%	Pred. No. 0.033			
	Matches 82	Conservative 0	Mismatches 73	Indels 0	Gaps 0	
QY	174 ggcatacgcgttaccctccggaagcggtgttggaaagacagacaccacgcgcaa	233				
Db	304 GCATGTAGTGAATGATAAGCTCAGGTAAAGCGGTGTGGTTAAAGCACCCACGAGTGGA	245				
OY	234 tgcgtctctctctcgcctcgttaccagttctcagttgctgcacattgacgcgcgacttgg	293				
Db	244 TTTAAGCATCGCTTTAGCGCAATTTAACCACAAAGTGGGGTTACTAGACGCTGATGTGTA	185				
OY	294 tctccgtaacctcgatctcctccttaggttagaga	328				
Db	184 TGGCCCTATATCCCTAGATGATGGGCTTGCAAA	150				
RESULT 11						
ID	Z32023/C					
ID	Z32023 standard; DNA; 30600 BP.					
AC	Z32023;					
DT	10-JAN-2000 (first entry)					
DE	Human METH1 related EST Z69361.					
XX	Human; METH1; METH2; anti-angiogenic; metalloproteinase thrombospondin;					
XX	cancer; diagnosis; hyperproliferative disorder; autoimmune disease;					
XX	angiogenesis inhibitor; abnormal wound healing; inflammation;					
XX	hematoid arthritis; psoriasis; endometrial bleeding disorder;					
XX	diabetic retinopathy; macula degeneration; haemangioma; detection;					
XX	arterial-venous malformation; immune deficiency; ss.					
OS	Homo sapiens.					
XX						
PN	WO9337660-A1.					
XX						
PD	29-JUL-1999.					
XX						
PF	22-JAN-1999; 99WO-US01313.					
XX						

PR 23-JAN-1998: 9805-0072298.  
 XX 28-AUG-1998: 9805-0098539.  
 XX  
 PA (TRUE/) IRUELA-ARISPE L.  
 PA (HST/) HASTINGS G A.  
 PA (RUBE/) RUBEN S M.  
 XX  
 PI Iruele-Arispe L, Hastings GA, Ruben SM;  
 XX  
 DR WPI: 1999-590684/50.  
 XX  
 PT New isolated metalloprotease thrombospondin polypeptides, useful for  
 PT treating hyperproliferative disorders, cancers or autoimmune disorders  
 PT  
 PS  
 PS Disclosure: Page 327-347; 457pp; English.  
 XX  
 CC 232000 and 232001 encode, and Y49501 and Y49502 represent, human  
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2  
 CC respectively. METH1 and METH2 have been found to be potent inhibitors of  
 CC angiogenesis both in vitro and in vivo. They can be used for treating  
 CC cancer and other disorders related to angiogenesis including abnormal  
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis,  
 CC endometrial bleeding disorders, diabetic retinopathy, some forms of  
 CC macular degeneration, haemangiomas, and arterial-venous malformations.  
 CC They may be useful in treating deficiencies or disorders of the immune  
 CC system, by activating or inhibiting the proliferation, differentiation,  
 CC or mobilisation (chemotaxis) of immune cells. The etiology of these  
 CC immune deficiencies or disorders may be genetic, somatic, such as  
 CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or  
 CC toxins), or infectious. They can also be used to treat inflammatory or  
 CC conditions, both chronic and acute conditions. The products can also be  
 CC used for detection and diagnosis. 232002 to 232080, and Y49503 to Y49511  
 CC represent sequences given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 30600 BP; 10390 A; 5313 C; 5429 G; 9467 T; 1 other:

Query Match 3.8%; Score 37.4; DB 20; Length 30600;  
 Best Local Similarity 50.2%; Pred. No. 0.37;  
 Matches 118; Conservative 0; Mismatches 116; Indels 1; Gaps 1;

QY 322 ttgaggaatcggatcgaattcgaattcgcgtgaggtataaaggagatgctgcgtcat 381  
 DB 1794 ttvcaaaacatgctcagatcagatcagatcagatcagatcagatcagatcagat 1735  
 QY 382 caagctcgtgacgtatgaatgaagcgttgctgaattcgaattcgtatatacctaaacct 441  
 DB 1734 gtatg-tgtttcttggatggatggatgctcctcaagaagatttcattgcttgcgaat 1676  
 QY 442 agatcgaatcctcgaatgagatgagatgagatgagatgagatgagatgagatgagatg 501  
 DB 1675 ggaaactgaagtgatgacatgcttggatgacatgcttggatgacatgcttggatgac 1616  
 QY 502 aaactgaagcgaaggttcacccgagatcattcatcattcattcattcattcattcatt 556  
 DB 1615 acatgttaattttagcaccattgcccatttttgagagaatgcatgattgagatgctgtaa 1561

RESULT 12  
 ID 236862  
 XX 236862 standard; DNA; 6659 BP.  
 AC  
 XX 236862;  
 XX  
 DT 13-MAR-2000 (first entry)  
 DE  
 XX DNA encoding a GON-1 protein of *Caenorhabditis elegans*.  
 DE  
 XX GON-1; metalloprotease; cell migration; modulator;  
 KW metalloprotease domain; thrombospondin domain; abnormal cell migration;  
 KW organ shaping; sterility; cancer metastasis; ss.

XX  
 OS *Caenorhabditis elegans*.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..6453  
 FT /\*tag= a  
 FT /product= "GON-1"  
 XX  
 PN MO961656-A2.  
 XX  
 PD 02-DEC-1999.  
 XX  
 PF 28-MAY-1999: 99MO-US11918.  
 XX  
 PR 29-MAY-1998: 9805-0087170.  
 PR 13-APR-1999: 9905-0129023.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 PI Kimble JE, Bleiloch RH;  
 DR WPI: 2000-072633/06.  
 DR E-PSDB: Y53898.  
 PT Identifying modulators of proteins containing metalloprotease and  
 PT thrombospondin domains, potentially useful for controlling cell  
 PT migration and organ shaping  
 XX  
 PS Disclosure: Page 39-51; 60pp; English.  
 XX  
 CC The present sequence encodes a GON-1 protein of *Caenorhabditis elegans*.  
 CC GON-1 is a secreted metalloprotease that lacks a transmembrane domain  
 CC and possesses a predicted metalloprotease domain between amino acids  
 CC 269-456. In *C. elegans* hermaphrodites, GON-1 is required for migration of  
 CC two distal tip cells to produce elongated tubes, whereas in males, GON-1  
 CC is required for migration of a single linker cell to produce a single  
 CC elongated tube. The protein is used in the method of the invention. The  
 CC specification describes a method for identifying a modulator of a  
 CC protein that contains a metalloprotease domain and a thrombospondin  
 CC domain. The method comprises treating a target organism, having a  
 CC developing gonadal cell that is responsive to the protein, with a test  
 CC compound, and determining any change in migration or shape of the cell  
 CC attributable to the test compound. The compounds identified are  
 CC potential therapeutic modulators of abnormal cell migration and organ  
 CC shaping, e.g. for rendering animals (specifically nematodes) sterile  
 CC and for inhibiting cancer metastases.  
 XX  
 SQ Sequence 6659 BP; 1928 A; 1337 C; 1547 G; 1847 T; 0 other:

Query Match 3.6%; Score 35.6; DB 21; Length 6659;  
 Best Local Similarity 50.0%; Pred. No. 0.61;  
 Matches 115; Conservative 0; Mismatches 114; Indels 1; Gaps 1;

QY 322 ttgaggaatcggatcgaattcgaattcgcgtgaggtataaaggagatgctgcgtcat 381  
 DB 3479 ttacaaacacatgctcagatgaatcattgccaataataaacttgaggaatggtctcagt 3538  
 QY 382 caagctcgtgacgtatgaatgaagcgttgctgaattcgaattcgtatatacctaaacct 441  
 DB 3539 gtatg-tgttcttggatggatggatgctgcgaagaaggttccatcattgcttgcgaat 3597  
 QY 442 agatcgaatcctcgaatgagatgagatgagatgagatgagatgagatgagatgagatg 501  
 DB 3598 ggaactgaagtcagatgacatgcttggatgacatgcttggatgacatgcttggatgac 3657  
 QY 502 aaactgaagcgaaggttcacccgagatcattcatcattcattcattcattcattcatt 551  
 DB 3658 acatgttaattaggaattgcccatttttgagagaatcattcattcattcattcattcatt 3707

RESULT 13  
 X07125

XX	ID	X07125 standard; DNA; 1590 BP.
XX	AC	X07125;
XX	DT	21-MAY-1999 (first entry)
XX	DE	Staphylococcus aureus mutant plic12 virulence gene.
XX	KM	Virulence: plic12; vaccine: antibacterial; antiphilic;
XX	KN	screening; bacteraemia; ss.
XX	OS	Staphylococcus aureus.
XX	PN	W09901473-A2.
XX	PD	14-JAN-1999.
XX	PF	03-JUL-1998; 98WO-GB01974.
XX	PR	03-JUL-1997; 97US-0887534.
XX	PA	(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX	PI	Holden DW;
XX	DR	WPI; 1999-105999/09.
XX	DR	P-PSDB; W97715.
XX	PT	Inhibition of virulence genes from Staphylococcus aureus - useful
XX	PT	for, e.g. screening for potential anti-microbial agents
XX	PS	Claim 1; Page 152-154; 203pp; English.
XX	CC	This is the nucleotide sequence of a virulence gene identified in
XX	CC	Staphylococcus aureus mutant plic12. S. aureus genes (see X07086-136)
XX	CC	associated with virulence were identified by signature-tagged
XX	CC	mutagenesis in which mutants containing a chromosomal insertion of
XX	CC	a signature tagged transposon were generated, mutants with
XX	CC	attenuated virulence were identified in a mouse model of bacteraemia,
XX	CC	and the nucleotide sequences of the regions flanking the transposon
XX	CC	insertion sites of these mutants were determined. Database sequence
XX	CC	comparisons were performed to identify the virulence genes and to
XX	CC	determine the possible function of their protein products (see also
XX	CC	W97680-724). The plic12 virulence gene product (see W97715) is
XX	CC	unknown, but has 40% identity to Mycoplasma mycoides hypothetical
XX	CC	protein in fth 5' region g01444. A claimed method of identifying
XX	CC	an antibacterial agent involves assaying potential agents for the
XX	CC	ability to interfere with the expression of S. aureus virulence
XX	CC	gene products. Also new is S. aureus containing a functional
XX	CC	mutation in one of the virulence genes, and its use in vaccine
XX	CC	compositions.
XX	Sequence	1590 BP; 617 A; 232 C; 340 G; 401 T; 0 other;

Query Match	Similarity	Score	DB	Length
Best Local	48.3%	Pred. No. 0.45;		
Matches	98;	Conservative	0;	Mismatches 105; Indels 0; Gaps 0;
QY	595	gaagcagcttcggttaacaactccggatatacaacgcgttaagagatgctatgggttaacg	654	
Db	778	gaagcggcttatatactctggtttgttccaaataagaagaatctcagaacaagactt	837	
QY	655	ggcttcctgaatgcgatgaalcaagaatataagaatgatgtgcagacagtggaact	714	
Db	838	gttaacttagtatctgtagtgagctatcatccagtgagatgaatgaatgatgtgcgaaaaa	897	
QY	715	gatatgattataaggaagaagatatgatgctcaagtttgatgctgcagagatggttgatg	774	
Db	898	gctaaagaagaatagacatatataatagaagaacagtggaacaacgtcatttgaaagtc	957	
QY	775	tcattgcttggtgtaattcctga	797	

Db	958	aacgcataatgcatcctga	980
RESULT	14		
ID	V74359	standard; DNA:	7778 BP.
AC	V74359;		
XX	16-MAR-1999	(first entry)	
DE	Staphylococcus aureus contig SEQ ID #48.		
XX	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KM	skin infection; surgical wound infection; scalded skin syndrome;		
KX	toxic shock syndrome; ds.		
OS	Staphylococcus aureus.		
XX			
FH	Key	Location/Qualifiers	
FT	misc_feature	1681..1740	
FT	/tag= a	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	3481..3540	
FT	/tag= b	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	5281..5340	
FT	/tag= c	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	7081..7140	
FT	/tag= d	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature		
PN	EP76519-A2.		
PD	30-JUL-1997.		
PF	07-JAN-1997;	97BP-0100117.	
PR	05-JAN-1996;	96US-0009861.	
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;		
PT	Rosen CA;		
WP	1997-374922/35.		
XX			
XX	Polynucleotide(s) and proteins derived from Staphylococcus aureus -		
CC	of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or		

FT	CDS	complement (441042..441899)
FT		/*tag= s
FT		/standard_name= "ORF k19"
FT		

```

FT /gene= "fixA"
FT /product= "protein required for nitrogenase activity"
FT complement (442316..442636)
FT /tag= t
FT /standard_name= "ORF K20"
FT /product= "protein of unknown function"
FT complement (443313..443879)
FT /tag= u
FT /standard_name= "ORF K21"
FT /product= "protein of unknown function"
FT 444337..445029
FT /tag= v
FT /standard_name= "ORF K22"
FT /product= "ferredoxin-like protein"
FT /note= "homologous to the NifH gene"
FT 445088..446602
FT /tag= w
FT /standard_name= "ORF K23"
FT /gene= "dctA"
FT /product= "C4-dicarboxylate transport protein"
FT /note= "homologous to the DctAI gene"
FT 446599..447843
FT /tag= x
FT /standard_name= "ORF L1"
FT /product= "cytochrome P450-like protein"
FT /note= "homologous to the CamC gene"
FT 447844..448500
FT /tag= y
FT /standard_name= "ORF L2"
FT /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT protein"
FT /note= "homologous to the LinA gene"
FT 448497..450203
FT /tag= z
FT /standard_name= "ORF L3"
FT /product= "putative protein with degradative function"
FT 450341..451396
FT /tag= aa
FT /standard_name= "ORF L4"
FT /product= "luciferase alpha-subunit-like protein"
FT /note= "homologous to the luxA gene"
FT 452980..454494
FT /tag= ab
FT /standard_name= "ORF L6"
FT /gene= "nifD"
FT /product= "alpha-subunit of Fmo protein of nitrogenase"
FT 454590..456131
FT /tag= ac
FT /standard_name= "ORF L7"
FT /gene= "nifK"
FT /product= "beta-subunit of Fmo protein of nitrogenase"
FT 456187..457677
FT /tag= ad
FT /standard_name= "ORF L8"
FT /product= "protein involved in Fmo co-factor
FT biosynthesis"
FT /note= "homologous to the NifB gene"
FT 457687..459096
FT /tag= ae
FT /standard_name= "ORF L9"
FT /product= "protein involved in Fmo co-factor
FT biosynthesis"
FT /note= "homologous to the FixF gene"
FT 459093..459575
FT /tag= af
FT /standard_name= "ORF L10"
FT /product= "protein of unknown function"
FT /note= "homologous to the NifX gene"
FT 459579..460067
FT /tag= ag
FT /standard_name= "ORF L11"
FT /product= "protein of unknown function"
FT 460501..460920

```

```

FT /tag= ah
FT /standard_name= "ORF L12"
FT /product= "protein similar to part of the Fe protein
FT of nitrogenase"
FT /note= "homologous to the NifH gene"
FT 461228..461545
FT /tag= ai
FT /standard_name= "ORF L13"
FT /product= "protein of unknown function"
FT 461301..464739
FT /tag= aj
FT /standard_name= "ORF L14"
FT /product= "peptidase-like protein"
FT /note= "homologous to the bi-MPP gene"
FT 464736..466079
FT /tag= ak
FT /standard_name= "ORF L15"
FT /product= "processing protease-like protein"
FT /note= "homologous to the pp gene"
FT 466590..467021
FT CDS

```

Query Match 3.5%; Score 34.6; DB 19; Length 534720;  
 Best Local Similarity 54.3%; Pred. No. 13; Gaps 0;  
 Matches 70; Conservative 0; Mismatches 59; Indels 0;

```

QY 161 ccgagaaacgcgcgtatcgttaccctccgaaagcggtgtgaaagcga 220
    |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 56482 CGGGGANAAGCTCCAGTCATTCGCCGTTGTCAATTTCAAGCGGTCGGAACACGA 56423

QY 221 caaccacgcgaagtcgtctctctcgtcgttaaggttccagttgctgcacatg 280
    |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 56422 CGACGGCCGCCCATCTCCGCGCATGACATGCGGTGACCGGACCGGTGCTTGCGGTGG 56363

QY 281 acgcgcgacc 289
    |||||
Db 56362 ACCTCGACC 56354

```

Search completed: January 8, 2001, 09:29:10  
 Job time: 20819 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2001, 03:38:45 ; Search time 81.49 Seconds

(without alignments)  
1934.161 Million cell updates/sec

Title: US-09-553-431-1

Perfect score: 978  
Sequence: 1 atggcgctctctgagatgttcttctctctcttcttgcggc 978

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2-6/ptodata/2/lna/5A\_COMB.seq: \*  
2: /cgn2-6/ptodata/2/lna/5B\_COMB.seq: \*  
3: /cgn2-6/ptodata/2/lna/6\_COMB.seq: \*  
4: /cgn2-6/ptodata/2/lna/PCRUS\_COMB.seq: \*  
5: /cgn2-6/ptodata/2/lna/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
C 1	46.6	4.8	7218	1	Sequence 14, Appl
C 2	39	4.0	1107	2	Sequence 1, Appl
C 3	34	3.5	7218	1	Sequence 14, Appl
C 4	33.6	3.4	44377	2	Sequence 7, Appl
C 5	33.6	3.4	44377	2	Sequence 1, Appl
C 6	32.2	3.3	1866	3	Sequence 13, Appl
C 7	31.4	3.2	72928	3	Sequence 1, Appl
C 8	31.2	3.2	1964	2	Sequence 8, Appl
C 9	30.2	3.1	1041	3	Sequence 4, Appl
C 10	29.6	3.0	4928	1	Sequence 1, Appl
C 11	29.6	3.0	5000	1	Sequence 5, Appl
C 12	29.6	3.0	246240	2	Sequence 20, Appl
C 13	29.6	3.0	246240	2	Sequence 21, Appl
C 14	29.6	3.0	246240	2	Sequence 22, Appl
C 15	29	3.0	2325	2	Sequence 5, Appl
C 16	29	3.0	2325	2	Sequence 5, Appl
C 17	29	3.0	2325	3	Sequence 5, Appl
C 18	29	3.0	3956	2	Sequence 9, Appl
C 19	29	3.0	3956	2	Sequence 9, Appl
C 20	29	3.0	3956	3	Sequence 9, Appl
C 21	28.8	2.9	740	3	Sequence 1, Appl
C 22	28.8	2.9	740	3	Sequence 1, Appl
C 23	28.8	2.9	3126	3	Sequence 3, Appl
C 24	28.6	2.9	433	1	Sequence 2, Appl
C 25	28.6	2.9	613	1	Sequence 1, Appl
C 26	28.6	2.9	2989	3	Sequence 10, Appl
C 27	28.6	2.9	3240	1	Sequence 34, Appl
C 28	28.6	2.9	3240	2	Sequence 34, Appl

C 29	28.6	2.9	3412	1	US-08-162-081B-32	Sequence 32, Appl
C 30	28.6	2.9	3412	2	US-08-780-872-32	Sequence 32, Appl
C 31	28.6	2.9	5565	3	US-09-068-195-21	Sequence 21, Appl
C 32	28.6	2.9	9468	1	US-08-325-547-10	Sequence 10, Appl
C 33	28.6	2.9	13987	2	US-08-804-227C-13	Sequence 13, Appl
C 34	28.6	2.9	43280	2	US-08-804-227C-13	Sequence 1, Appl
C 35	28.6	2.9	1215	4	PCT-US96-05320A-707	Sequence 707, App
C 36	28.4	2.9	2101	2	US-08-637-763B-5	Sequence 5, Appl
C 37	28.4	2.9	2101	3	US-09-170-354-5	Sequence 5, Appl
C 38	28.4	2.9	9045	3	US-09-121-321-1	Sequence 1, Appl
C 39	28.2	2.9	928	5	5171685-3	Patent No. 5171685
C 40	28.2	2.9	928	5	5171685-3	Patent No. 5171685
C 41	28.2	2.9	956	1	US-08-624-125-3	Sequence 3, Appl
C 42	28.2	2.9	1838	3	US-08-470-260-7	Sequence 7, Appl
C 43	28.2	2.9	1838	3	US-08-471-491-7	Sequence 7, Appl
C 44	28.2	2.9	1838	3	US-08-466-662-7	Sequence 7, Appl
C 45	28.2	2.9	3306	1	US-08-261-206A-71	Sequence 71, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHIEFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14  
Query Match 4.8%; Score 46.6; DB 1; Length 7218;

# TELECOMMUNICATION INFORMATION:

us-09-553-431-1.rni

US-08 05  
US-08-804-198-1  
Sequence 1, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kunitz, Stuart A.  
APPLICANT: Rao, Nagatja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rosteck, Paul R., Jr.  
TITLE OF INVENTION: PLATENOLOIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL, 1138  
STREET, LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1

```

; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -

```

QY 378 cgaatcaagcctcgttaagtatagaagcttgctgaatttcgaatt 422  
 Db 49154 GTAAATGCGTATGCGCCGGGGAAGAGAGAGATTCAGATT 49198

RESULT 8  
US-08-841-349-8/c  
Sequence 8, Application US/08841349B  
Patent No. 5955594  
GENERAL INFORMATION:  
APPLICANT: MISHRA, LODA  
TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.  
FILE REFERENCE: XX/PO4470US0  
CURRENT APPLICATION NUMBER: US/08/841,349B  
CURRENT FILING DATE: 1997-04-30  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1964  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (322)..(1509)  
US-08-841-349-8

Query Match 3.2%; Score 31.2; DB 2; Length 1964;  
Best Local Similarity 52.3%; Pred. No. 2.1;  
Matches 69; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 409 tcgaattcgaattcgtatgtatcctaaacctgaacatccgatgagattgt 468  
DB 207 TAGCGTTCATATAATAATAATACATTCATTAATAATAATCAATAGTTCG 148  
QY 469 gctaaagcatggaatgagctgtgagtcgttgaacacagaccggaagctaccgagat 528  
DB 147 TGGAAAGGTTGGATCTTTTGTGCGATGATACATGCGTAAAGCATATTAATCT 88  
QY 529 ttcatcatcattc 540  
DB 87 TAAATCATCTTC 76

RESULT 9  
US-09-033-055A-4  
Sequence 4, Application US/09033055A  
Patent No. 6069241  
GENERAL INFORMATION:  
APPLICANT: OHKAWA, HIDEO  
APPLICANT: IMAISHI, HIROMASA  
TITLE OF INVENTION: NOVEL CYTOCHROME P450 GENE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY, MADISON & SUTRO  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/033,055A  
FILING DATE: MARCH 2, 1998  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: PAUL E. WHITE, JR.  
REGISTRATION NUMBER: 32,011  
REFERENCE/DOCKET NUMBER: 9437/251563  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3651  
TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1041  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-033-055A-4

Query Match 3.1%; Score 30.2; DB 3; Length 1041;  
Best Local Similarity 46.8%; Pred. No. 3.1;  
Matches 95; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 649 gctaaagggttctagaatgcagatggaatcaagatataaagatgctgaaacagatg 708  
DB 266 GATAGCGTGCAGAAATATGAGATTCACAGCATGATGAGATGATGAGAAAGTG 325  
QY 709 agaatgatataatgaagagagagatataatgctagtgatgctcaggaagatgtg 768  
DB 326 AAATTGGGAATTAATGAGAGAGTAAAGTGAATTGGAGAAATGCTGAGAGAAAGT 385  
QY 769 ggaatgcatctgtgtgtaattcctgaagatctcgaaggtatcgaagccagatcga 828  
DB 386 GAATATATGGGAATGCGAGAGAAAGCAAAATGATAGTCGGGGAATATATAAGAAATGTA 445  
QY 829 ggaatcgcgtgtgtcgaataa 851  
DB 446 AGGGCTGTGTATGTTATGGAAA 468

RESULT 10  
US-08-399-561-1  
Sequence 1, Application US/08399561  
Patent No. 5752903  
GENERAL INFORMATION:  
APPLICANT: Hirschberg, Joseph  
APPLICANT: Cunningham Jr., Francis X.  
APPLICANT: Gant, Elisabeth  
TITLE OF INVENTION: Lycopene Cyclase Gene  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kohn & Associates  
STREET: 30500 No. 5792903thwestern Highway, Suite 410  
CITY: Farmington Hills  
STATE: MI  
COUNTRY: US  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/399,561  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth I.  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: P-305 (Hebrew Univ.)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 810-539-5050  
TELEFAX: 810-539-5055  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4928 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS

LOCATION: 2029..3261  
US-08-399-561-1

Query Match 3.0%; Score 29.6; DB 1; Length 4928;  
Best Local Similarity 52.4%; Pred. No. 12;  
Matches 89; Conservative 0; Mismatches 79; Indels 2; Gaps 1;

QY 783 tgggttaattcctgaagattcgaaggttttccgaacacgaaggttccgctgt 842  
DB 1663 TGCTGCTCAGCGTCACATTCGCCGCCGTACCGAAGAGGAATCGCTGCCTGA 1722  
QY 843 tctgaataagcctcgaagcttcgaggttgcgttgcagagcgcttgagactcgt 902  
DB 1723 CCGGCCAAGGCGCTCGACGCTACGTCGAT--CTGTTGACACCGGCTCATCCGCGG 1780  
QY 903 tgaagaagaatgatagaagcgctgtatgttgagaggaagaccataagaac 952  
DB 1781 TAAGCCCTATTCTCTGCTTAGCGGTAGGTGCTGTAAGAAAGTGAACCCAC 1830

## RESULT 11

US-09-147-522-5/C  
Sequence 5, Application US/09147522  
Patent No. 6107069  
GENERAL INFORMATION:  
APPLICANT: MAGAGNIN, SIMONA  
APPLICANT: BENATTI, LUCA  
APPLICANT: CINI, MASSIMO  
APPLICANT: SPECIALE, CARMELA  
APPLICANT: COVINI, NEVIE  
TITLE OF INVENTION: RECOMBINANT KYURENINE-3-HYDROXYLASE ENZYME AND  
FILE REFERENCE: 0769-0408-0PCT  
CURRENT APPLICATION NUMBER: US/09-147,522  
EARLIER FILING DATE: 1999-01-14  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 5  
LENGTH: 5000  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (47)..(1507)  
US-09-147-522-5

Query Match 3.0%; Score 29.6; DB 3; Length 5000;  
Best Local Similarity 53.4%; Pred. No. 13;  
Matches 62; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 601 gttctgtaaacactccgatatcaacgcttaagagagctgatagggttacgggttg 660  
DB 4309 GTGAGATTGAAGCTTGAGAGGATTACAAAGTATTGAATGCTTCAATATTCGGTTAA 4250  
QY 661 ttgaatgcgattgagatcagaagataaagatgattgtgaacagagtgagactga 716  
DB 4249 TAAGATTAGATGAATTAAGACATTAAGCTCTGGGCTGATTAGCTAGATCTGA 4194

## RESULT 12

US-08-724-394A-20  
Sequence 20, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Laufer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereo  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fites, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note="HLA-H.CONTIG"  
US-08-724-394A-20

Query Match 3.0%; Score 29.6; DB 2; Length 246240;  
Best Local Similarity 52.4%; Pred. No. 1.2e+02;  
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 40 cttctcctcatatcctctcacaagaactcaatatacttcacgaagattcgtcaat 99  
DB 152799 CATCTTCACTAACTTCTCTTCTACTCAAAATTCCAAAATATTTTCATCAGGATTTACAT 152858  
QY 100 aacctagcagcgaagtcacatcagatcgcgttctcaatttaacgcgaacgcgaactc 159  
DB 152859 AATAGACAAATGGTGATTTTAAATATAAAGAGAGATTGATGATAATAGCAGGCTGACCA 152918  
QY 160 gccg 163  
DB 152919 TCCG 152922

## RESULT 13

US-08-724-394A-21  
Sequence 21, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Laufer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereo

RESULT 14  
US-08-724-394A-22  
Sequence 22, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Laufer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchinashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1  
TITLE OF INVENTION: Sequences and Antibodies Thereeto  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP  
STREET: Two Embarcadero Center, 8th Floor

```

Query Match      3.0%; Score 29.6; DB 2      Length 246240;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0.

QY  40  ctctctctcatcatcctctctctcaaaaagactcaatctctccacaagatctgcacat 99
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152799 CATCTTCAACTAAACTTTCTCTTACTCAAAATTTCCAAATATTTTTCATCAGGATTTCTAAACAT 152858

QY  100 aaccttagcagacggagctccaatcagctacgctcgtctcttcaattcaatcgcnaaacgcgaactc 159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152859 AATCAGACAATGCGTATTTAATAAAAAAGAGAGATTTGAGTGAATAATAGCAGCGCTGAACCA 152918

QY  160 gccgc 163
      |||
Db 152919 TCCG 152922

```

15  
 US-08-966-388-5  
 Sequence 5, Application US/08966388  
 Patent No. 5965412  
 GENERAL INFORMATION:  
 APPLICANT: TOMOYUKI NISHIMOTO  
 APPLICANT: MICHIO KUBOTA  
 APPLICANT: HIROTO CHAEN  
 APPLICANT: TOSHIO MIYAKE  
 TITLE OF INVENTION: KOUJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSSEE: BROWDY AND NEIMARK  
 STREET: 419 Seventh Street, N.W., Suite 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/966,388  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 311,235/1996  
FILING DATE: 8-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

Query Match	3.0%	Score 29	DB 2	Length 2325
Best Local Similarity	50.4%	Pred No. 13		
Matches 71, Conservative	0	Mismatches 70	Indels 0	Gaps 0

Qy-	551	gaatgaaccaaattctcgtgaacaactcccgagataaacagcgttlaaggagatcatalaaggt	650
Db	267	GTATGTTGAGGCTGAGAGATTAAAGTTGGATTAATGCCAGTTGATAGATTAAAGCAAT	326
Qy	651	taacggttctgttaagaatbcgaatgaaatcagagataaagaatgattcttgaacaagatgag	710
Db	327	TCTTGATATTGAAAAAGGATATTCTATTATTAGAGATATGTTGATTAAAGACAGCAAAAGATTAG	386
Qy	711	aactgatgatgattlaaaggaga	731
Db	387	AATTACTGCAATTGAGGCATA	407

Search completed: January 8, 2001, 09:20:42  
Job time: 20517 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2001, 08:31:03 ; Search time 1361.86 Seconds

(without alignments)  
6082.013 Million cell updates/sec

Title: US-09-553-431-3

Perfect score: 1182

Sequence: 1 aaagctgataatcgcaactcc.....tcaaaaaaaaaaaaaaaaaa 1182

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues  
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estbda:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_esta1:\*  
88: gb\_esta2:\*  
89: gb\_esta3:\*  
90: gb\_esta4:\*  
91: gb\_esta5:\*  
92: gb\_esta6:\*  
93: gb\_esta7:\*  
94: gb\_esta8:\*  
95: gb\_esta9:\*  
96: gb\_esta10:\*  
97: gb\_esta11:\*  
98: gb\_esta12:\*  
99: gb\_esta13:\*  
100: gb\_esta14:\*  
101: gb\_esta15:\*  
102: gb\_esta16:\*  
103: gb\_esta17:\*  
104: gb\_esta18:\*  
105: gb\_esta19:\*  
106: gb\_esta20:\*  
107: gb\_esta21:\*  
108: gb\_esta22:\*  
109: gb\_esta23:\*  
110: gb\_esta24:\*  
111: gb\_esta25:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: em\_estp25:\*  
 127: em\_estp26:\*  
 128: em\_estp27:\*  
 129: em\_estp28:\*  
 130: em\_estp29:\*  
 131: em\_estp30:\*  
 132: em\_estp31:\*  
 133: em\_estp32:\*  
 134: em\_estp33:\*  
 135: em\_estp34:\*  
 136: em\_estp35:\*  
 137: em\_estp36:\*  
 138: em\_estp37:\*  
 139: em\_estp38:\*  
 140: em\_estp39:\*  
 141: em\_estp40:\*  
 142: em\_estp41:\*  
 143: em\_estp42:\*  
 144: em\_estp43:\*  
 145: em\_estp44:\*  
 146: em\_estp45:\*  
 147: em\_estp46:\*  
 148: em\_estp47:\*  
 149: em\_estp48:\*  
 150: em\_estp49:\*  
 151: em\_estp50:\*  
 152: em\_estp51:\*  
 153: em\_estp52:\*  
 154: em\_estp53:\*  
 155: em\_estp54:\*  
 156: em\_estp55:\*  
 157: em\_estp56:\*  
 158: em\_estp57:\*  
 159: em\_estp58:\*  
 160: em\_estp59:\*  
 161: em\_estp60:\*  
 162: em\_estp61:\*  
 163: em\_estp62:\*  
 164: em\_estp63:\*  
 165: em\_estp64:\*  
 166: em\_estp65:\*  
 167: em\_estp66:\*  
 168: em\_estp67:\*  
 169: em\_estp68:\*  
 170: em\_estp69:\*  
 171: em\_estp70:\*  
 172: em\_estp71:\*  
 173: em\_estp72:\*  
 174: em\_estp73:\*  
 175: em\_estp74:\*  
 176: em\_estp75:\*  
 177: em\_estp76:\*  
 178: em\_estp77:\*  
 179: em\_estp78:\*  
 180: em\_estp79:\*  
 181: em\_estp80:\*  
 182: em\_estp81:\*  
 183: em\_estp82:\*  
 184: em\_estp83:\*  
 185: em\_estp84:\*  
 186: em\_estp85:\*  
 187: em\_estp86:\*  
 188: em\_estp87:\*  
 189: em\_estp88:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	295	25.0	710	93	AM668240	AM668240 GA_Fa001
2	253.6	21.5	873	94	AM729746	AM729746 GA_Fa002
3	190	16.1	557	107	BE35432	BE35432 DGI_40_A0
4	187.6	15.9	456	28	AL370067	AL370067 MCB35D01
5	131	11.1	588	107	BE359345	BE359345 DGI_40_A0
6	129.8	11.0	773	108	BE418872	BE418872 SCL083.D0
7	126.2	10.7	278	40	AM160136	AM160136 pblt12 so
8	122.6	10.4	276	174	B23006	B23006 F16H1TF IGF
9	122	10.3	362	104	BE124676	BE124676 EST9371I
10	97.2	8.2	457	140	C99848	C99848 C99848 YAC
11	73.4	6.2	386	92	AM574469	AM574469 707092H09
12	72.8	6.2	1101	190	CNS0039C	AL063921 Drosophila
13	61.6	5.2	1146	191	CNS021G2	AL076843 Tetradon
14	58	4.9	1101	190	CNS00LT2	AL078714 Drosophila
15	57.8	4.9	1101	190	CNS000D1	AL065414 Drosophila
16	57.2	4.8	1101	190	CNS00LOO	AL068607 Drosophila
17	57	4.8	1101	190	CNS0100X	AL098379 Drosophila
18	55.2	4.7	919	190	CNS005RL	AL061409 Drosophila
19	54.8	4.6	767	190	CNS00A0X	AL055924 Drosophila
20	53	4.5	1031	190	CNS00CF2	AL059199 Drosophila
21	52	4.4	895	190	CNS0071A	AL066286 Drosophila
22	51.2	4.3	622	29	AU060989	AU060989 AU060989
23	51.2	4.3	1100	190	CNS0161V	AL106189 Drosophila
24	51	4.3	878	190	CNS0187R	AL108993 Drosophila
25	51	4.3	1080	190	CNS00EPP	AL069494 Drosophila
26	51	4.3	1101	190	CNS00396	AL063911 Drosophila
27	50.4	4.3	402	36	AV408561	AV408561 AV408561
28	50.4	4.3	420	36	AV411765	AV411765 AV411765
29	50.4	4.3	422	36	AV411541	AV411541 AV411541
30	50.4	4.3	859	190	CNS0044Y	AL055406 Drosophila
31	50.2	4.2	1101	190	CNS00399	AL063914 Drosophila
32	50.2	4.2	1101	190	CNS0106X	AL098595 Drosophila
33	50.2	4.2	1101	190	CNS017KX	AL108171 Drosophila
34	50.2	4.2	1201	190	CNS016CN	AL106577 Drosophila
35	50	4.2	1101	190	CNS016XR	AL107337 Drosophila
36	50	4.2	1201	190	CNS015YI	AL106068 Drosophila
37	49.8	4.2	884	190	CNS006UO	AL065923 Drosophila
38	49.6	4.2	1204	190	CNS016E2	AL106628 Drosophila
39	49.4	4.2	187	39	AM085373	AM085373 wy62H02.x
40	49.4	4.2	1084	190	CNS00ZS5	AL097258 Drosophila
41	49.2	4.2	1003	190	CNS017UQ	AL108524 Drosophila
42	49.2	4.2	1201	190	CNS0107R	AL098625 Drosophila
43	49	4.1	939	190	CNS00CKG	AL059400 Drosophila
44	48.8	4.1	928	190	CNS00DKY	AL071865 Drosophila
45	48.8	4.1	1101	190	CNS00LXB	AL078903 Drosophila

## ALIGNMENTS

RESULT 1  
 LOCUS AM668240  
 DEFINITION GA\_Fa0013E22 Gossypium arboreum 7-10 dba fiber library Gossypium  
 ACCESSION AM668240  
 VERSION AM668240.1 GI:7502620  
 KEYWORDS EST.  
 SOURCE Gossypium arboreum.  
 06-APR-2000

ORGANISM *Gossypium arboreum*

REFERENCE

REFERENCE AUTHORS TITLE
1 (pages 1 to 710) Leslie, A., Fritsch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A. An integrated analysis of the genetics, development, and evolution

JOURNAL Unpublished (2000)  
COMMENT Contact: Wing RA

FEATURES	Location/Qualifiers
source	1. .710

BASE COUNT	185	a	/nole=vector:pbk-cmv; site_1: EcORI; site_2: XhoI
ORIGIN	157	c	
	213	g	
	155	t	

Query Match	25.0%;	Score 295;	DB 93;	Length 710;
Best Local Similarity	75.5%;	Pred. No. 3.6e-64;		
Matches 380; Conservative	0;	Mismatches 120;	Indels 3;	Gaps 1

OY	230	gtcgcgcgaatcgagccgacgcttgctcaagttaacccgcgaactctctcgtgctcgaa	289
Db	207	gTAGTGGCCATGAGCGCGAGCTTGCCCTCCGTTAACCTGAGACTCTCTTAAAGCTTGAG	266
OY	290	aaccgcgcgaattacaacgctgcttgtaagttcttcaacgcgcgalttcgaagctcgaccaagcc	349
Db	267	AATCCCGTGAATATACCGTCTGTGAGGAGTACATTAATGGGATTTGGCCGATGGACCGAGCA	326
OY	350	cbaagtcgtgaaacgcgcgtgtcaaatcttcgaattgcgtctgtatcttcaaaactagtgca	409
Db	327	TTTGCTAAAGAGTAACCGATGCTGCAACTTTCGAATTTGCTTCATTTCCAAAGCCAGATCG	386
OY	410	aaatgcgctttaggaattctgggggaaaagccttagtttg --- cttgagtcattaagaat	466
Db	387	AAGCTCCCCATTTGGATTGGTGGTGGAACCGTTGGCTTGTGTAGTGGACCGACTGGAAGCG	446
OY	467	aggcgaagaagcttgcgcggaatttctacttaagaatttgcctgcgaagtatctgacgg	526
Db	447	CGTAGGAAGAGGTCCCGGATTTCTACTAATAGATTGCTCCGGCCGGGATTGATCCAGCG	506
OY	527	ttcataaacgcgcaltacacgcggtctaaagaaagcgatattagttacacaacactgatalact	586
Db	507	TTTTCATAAGCGCAATAAACCCAGCAAAACGACGGCGGTTTTGGTGACGACCCCGGACATTACG	566
OY	587	gcaltgagaaga tgcagatagaagcttaacgctgtctgtgaattgtaattggaatttggaatc	646
Db	567	AGTTTAAGGGGATGCGGATAGAGTGCACAGCGGCTGTAGAAATGTGANTGGGATTAAGGATATA	626
OY	647	aaaaattcttggaacacaggtttagaactctatttgataaagggtgtgaagatatgattcagct	706
Db	627	AAGATGATAGTAAACACAGCTCAGCAGCGATATGATATAAGGAGAGAGATATGATGTCTGGTT	686
OY	707	cttgatgttcaagaagatgttggg 729	
Db	687	TTTGATGTGCAAGAAATGCTCTGGG 709	

## RESULT 2

AW729746

### DEFINITION

DEFINITION	ACCESSTION
GA_Ea0025013 Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA_Ea0025013, mRNA sequence.	AW220746

VERSION	AW/29/46.1
KEYWORDS	EST.
SOURCE	Gasevium arborum

ORGANISM

REFERENCE  
1. Chaceff, I. 1972. 111.

AUTHORS	TITLE
Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wling, R. A. and Wilkins, T. A.	An integrated analysis of the genetics, development, and evolution

JOURNAL Unpublished (2000)  
COMMENT Contact: Wing RA

FEATURES	Location/Qualifiers
source	1. .873

```

/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/cloneme="GA_E80025013"
/cloneme_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      216 a      186 c      268 g      203 t
ORIGIN

```

Query Match	21.5%;	Score 253.6;	DB 94;	Length 873;
Best Local Similarity	72.7%;	Pred. No. 1.1e-53;		
Matches 355; Conservative	0;	Mismatches 129;	Indels 4;	Gaps 2

[illegible]

Db	567	AAGTCTAATGATGGCTGTATTAAGTACAGACACGGCGTGTAGAAATGTGATGGATTAAGCATAT	626
Oy	646	taaatgctcttgcacagagttgatacgtatgttgataagggtgataagatatgatctcgt	705
Db	627	AAAGATGTGGGAAACAGATGTAGAGACCGCTCTGTATTAGAGGGGAAAGATATGATCCGCT	686
Oy	706	ctctgagt	713
Db	687	TCTGTATG	694
RESULT	3		
BE359432			
LOCUS	BE359432	557 bp	mRNA
DEFINITION	DG1_40_A03.g1_A002	Dark Grown 1 (DG1)	Sorghum bicolor cDNA, mRNA
ACCESSION	BE359432		
VERSION	BE359432		
KEYWORDS	EST.		
SOURCE	BE359432.1	GI:9300989	
ORGANISM	Sorghum bicolor		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.		
REFERENCE	1 (bases 1 to 557)		
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.		
TITLE	An EST database from Sorghum: dark-grown seedlings		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu		
	Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.		
	Seq primer: Polymix		
	High quality sequence start: 6		
	High quality sequence stop: 531		
	POLYA=No.		
FEATURES			
source	Location/Qualifiers		
	1..557		
	/organism="Sorghum bicolor"		
	/db_xref="taxon:4558"		
	/clone_lib="Dark Grown 1 (DG1)"		
	/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced by mass excision."		
BASE COUNT	116 a	127 c	174 g
ORIGIN		140 t	
Query Match	16.1%;	Score 190;	DB 107; Length 557;
Best Local Similarity	69.2%;	Pred. No. 1.2e-37;	
Matches	288;	Conservative 0;	Mismatches 125; Indels 3; Gaps 2
Oy	482	ccgagattatatactaatagatgttcctcgcagagatattgatgcccgggttcataacgcccaatt	541
Db	1	CCGCGCTTATCCTCATCGACACTGCCCGCAGAGGTGTGATGCCGGTGTTCACATGCCAATT	60
Oy	542	acaaccgctcaagaagccgattatagttaaaaccccgatatattctcgtatgagagatgca	601
Db	61	GCACCTGCAGAAAGACGACAGTGGCTGTTACACCCCTGACATTCAGGCTCTCCGTGATGCT	120
Oy	602	gataaggtttacaagctctgcttgaatttgatggaattaggagatattaaatgattgtgaac	661
Db	121	GACCGTGTGGCAGACACTGTTGGAGTCGAGATGCATCAAGATATATCAAGATTATTGTCAAC	180
Oy	662	agaagttagaactgattatgaagggtgaaagatatatgctcagttcttgatgltcaagag	721

Db	181	CGAATGGGGCCAGACACTGTCGTAAGGGGAGGACATGATGTCAGACACTGATGTCACAAAGA	240
QY	722	atgttgatgtatgtcatgtg--ttgagtgtatccagagatgc--gaagtatccgagatgcg	778
Db	241	ATGCTTGGGTTGGCCCTTCTGTTGGTGTGGTCTCTGAGATGGGAGATATCCGGAGTACAA	300
QY	779	aataagaggtttccagcttgatgtltaacaagctccgaacttagcagagatgtgcatltag	838
Db	301	AATAGGGGTGTACCACTTGCTGCTCAGACGACCGCGCCAGCGCGGCTTGTCTGGAC	360
QY	839	caggctgcttggagagatgtgttgagcaagatgcatgaaagctgtgatgtgttgaggga	894
Db	361	CAGGCTACTTGCGCGATTTGGTGGAAAGAGATGTGATGACAGCAGCATGATGATGAGA	416
RESULT	4		
LOCUS	AL370067	456 bp	mRNA
DEFINITION	MLBA35D01.F1	MLBA Medicago truncatula cDNA clone MLBA35D01 T3,	EST
ACCESSION	AL370067		03-AUG-2000
VERSION	AL370067		
KEYWORDS	AL370067.1	GI:9669820	
SOURCE	ESF		
ORGANISM	barrel medic.		
	Medicago truncatula		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;		
	Fabales; Fabaceae; Papilionoideae; Medicago.		
REFERENCE	1 (bases 1 to 456)		
AUTHORS	Journet,E.P., Crespeau,H., van-Tulden,D., Gouzy,J., Jailion,O.,		
	Niebel,A., Carreau,V., Chataignier,O., Kohn,D., Glaninazzl-Pearson		
	,V. and Gamas,P.		
	Medicago truncatula ESTs from nitrogen-starved roots		
TITLE	Medicago truncatula ESTs from nitrogen-starved roots		
JOURNAL	unpublished (2000)		
COMMENT	Genoscope - Centre National de Sequencage		
	Contact: Genoscope		
	Bp 191 91006 EVRY cedex - France		
	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Contact : Pascal Gamas and Etienne-Pascal Journet, laboratoire de		
	Biologie Molculaire des Relations Plantes-Microorganismes,		
	CNRS-INRA, Bp 27 31326 Castanet-Tolosan Cedex, France (Email :		
	Mt-est@toulouse.inra.fr Website :		
	http://sequence.toulouse.inra.fr/Mtruncatula.html).		
FEATURES	Location/Qualifiers		
Source	1..456		
	/organism="Medicago truncatula"		
	/cultivar="Jemalong"		
	/db_xref="taxon:3880"		
	/clone="MLBA35D01"		
	/clone_1ib="MEBA"		
	/tissue_type="root tips"		
	/dev_stage="harvested after 3 days of N-starvation"		
	/note="vector: phluescript psk; Site.1: EcorI; Site.2:		
	XhoI; Plants were grown in an aeroponic chamber for 14		
	days on nitrogen-rich medium followed by 3 days on N-free		
	medium. RNA was extracted from root tips (1-3 cm). cDNA		
	was prepared from polyA+ enriched RNA. The cDNA was		
	directionally ligated into uni-zapR vector from		
	Stratagene and packaged using giga-pack Gold packaging		
	extracts. Plasmids containing cDNA inserts were		
	mass-excised from phage stocks using ExsacII helper phage		
	and propagated in SOLR cells. Clone ordering and		
	sequencing was performed by the Centre National de		
	Sequencage (Genoscope, Evry, France)."		
BASE COUNT	110 a	68 c	141 g
ORIGIN	137 t		
Query Match	15.9%;	Score 187.6;	DB 28; Length 456;
Best Local Similarity	81.6%;	Pred. No. 4,6e-37;	
Matches 217; Conservative	0;	Mismatches 49;	Indels 0; Gaps 0
QY	476	gattgccggattataactatagatttccgcagagatattgatgcgggttcataaac	535

Db 188 GCGTCGCGACGCTTATCTGATGATGCGCTGACGAGGATGATGCGGGTTATTAACG 247  
 Oy 536 gcatltaacccggtcgaagaacgcgtatagtltaaacacctatattactgcatlaga 595  
 Db 248 GCGATTACACCGGCTAATATAGCGGTTTGCATATACACGCGGATATTAACCTGAGG 307  
 Oy 536 gctcgaatagatgaagctctgctgaatgcatgagaaataggaatataaatgatt 655  
 Db 308 GATGCGGATAGAGTGAAGGATGTTGAGATGATGATGATGATGATGATGATGATGATG 367  
 Oy 656 gtagaagaagcttagaagctatgataaggggtgaagatgtagtctgctgattgatt 715  
 Db 368 GTCATAGGCTTATGAGCAGGATATGATTAAGGATGATGATGATGATGATGATGATG 427  
 Oy 716 caaagaatgctggaatgcatgctt 741  
 Db 428 CAAGACATCTTGCGTTTGCGCTTGCCT 453

RESULT 5  
 BE359345 588 bp mRNA EST 20-JUL-2000  
 LOCUS DGL\_40\_A03\_b1\_A002 Dark Grown 1 (DGL) Sorghum bicolor cDNA, mRNA  
 DEFINITION BE359345  
 BE359345  
 VERSION BE359345.1 GI:9300902  
 KEYWORDS EST.  
 SOURCE Sorghum.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.  
 REFERENCE 1 (bases 1 to 588)  
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 L.H.  
 TITLE An EST database from Sorghum: dark-grown seedlings  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM,  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Pured quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: JEN REV  
 High quality sequence stop: 447  
 POLYA-NO.

FEATURES  
 source  
 1. 588  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGL)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 lambda zap; Site 1: XhoI; Site 2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda ZAP II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 77 a 264 c 160 g 87 t  
 ORIGIN

Query Match 11.1%; Score 131; DB 107; Length 588;  
 Best Local Similarity 58.7%; Pred. No. 9,7e-23;  
 Matches 227; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Oy 193 caatcgcaaacagagctgcgcgagagacatccacgagctgcgaatcgacgcgaagt 252  
 Db 201 CAACCTGCGCGCTCGCTCGCGCGCTCGGACTCCCGCGCTCGCGCTCGGAGCGCGCGC 260  
 Oy 253 tggctacgtaacctgactctcttcgctcgatgaagaacgcgcaataacagctgct 312

Db 261 CGGCTCCGCAACCTTCGACCTTCGCTGCGCCTCGAGAACCCGCGTCCACCTCACCGCCG 320  
 Oy 313 tgaagcttcaacggcgatgtagactgagaccgaagccctagttcgtgataaagctgctc 372  
 Db 321 CGACTCTCTCGTGGGAGACTGACGCTGACAGCGGCTGCTCGACACCGCGGCTCCA 380  
 Oy 373 aaattcgaaatgcttctgatttcaaaacatagatgcaaatgctttagattgagg 432  
 Db 381 GGACTCCACCTTCCTGCTGCTCCAGGCTCCAGGCTCCAGGCTCCAGGCTCCAGGCT 440  
 Oy 433 aaagcttagattgctgctgcatgcatlaaagaatagcaagaaggttgcgcggaatttat 492  
 Db 441 CAAGACCTCACCTGAGGTGCGGACCGCCGCTTCGCGCTCCCTCAACACCGCCGCTTCAT 500  
 Oy 493 actatagattgctcgcgagatgattgagtcgaggttcatcaaacgcgcatlaacgcgctaa 552  
 Db 501 CCTATCGACTGCGCGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 560  
 Oy 553 cgaagcgatgattgatacaacactga 579  
 Db 561 AGAAGAGTCTGCTGTTACCAACCCCTGA 587

RESULT 6  
 BE18872 773 bp mRNA EST 24-JUL-2000  
 LOCUS SCL083.D06R990805 ITFC SCL Wheat Leaf Library Triticum aestivum  
 DEFINITION BE18872  
 BE18872  
 VERSION BE18872.1 GI:9416718  
 KEYWORDS EST.  
 SOURCE bread wheat.  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
 REFERENCE 1 (bases 1 to 773)  
 Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,  
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,  
 Herrmann,R.G., Holton,T., Jacquemlin,J.M., Jia,J., Joudrier,P.,  
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,  
 Pecchioni,N., Quisenberry,C., Schuch,W., Selvaraj,G., Shariflou,M.,  
 Sorrells,M., Warburton,M. and Wenzel,G.  
 TITLE International Triticaceae EST Cooperative (ITFC): Production of  
 Expressed Sequence Tags for Species of the Triticaceae  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cloutier S  
 Cereals Research Centre, Agriculture & Agri-Foods Canada  
 Winnipeg MT CANADA  
 Tel: 204 983 2340  
 Fax: 204 983 4604  
 Email: scloutier@em.agr.ca  
 International Triticaceae EST Cooperative (ITFC)  
 http://wheat.pw.usda.gov/genome.  
 Location/Qualifiers

FEATURES  
 source  
 1. 773  
 /organism="Triticum aestivum"  
 /cultivar="Patcher Lrl"  
 /db\_xref="taxon:4565"  
 /clone\_lib="SCL083.D06"  
 /note="SCL083.D06"  
 /clone\_lib="ITFC SCL wheat Leaf Library"  
 /issue\_type="young leaf"  
 /dev\_stage="after 24 hour challenge with LR-AVR1"  
 /note="Vector: Lambda ZAP; 1.0 kbp average insert size."  
 BASE COUNT 117 a 278 c 250 g 115 t 13 others  
 ORIGIN

Query Match 11.0%; Score 129.8; DB 108; Length 773;  
 Best Local Similarity 59.7%; Pred. No. 2e-22;  
 Matches 252; Conservative 0; Mismatches 168; Indels 2; Gaps 2;

Oy 230 gtgctgcaatcgacgcgcgctgtgtgtctgatacctgactctcttcgctgctcga 289

Query Match	10.78; Score 126.2; DB 40; Length 278;
-------------	--

Query Match	10.4%	Score 122.6	DB 174	Length 276
Best Local Similarity	75.6%	Pred. No. 1.1e-20		
Matches 153	Conservative 0	Mismatches 49	Indels 0	Gaps 0
QY 754	agattcgaagtgcaggtacgaatagggtttccgcttggttgaacagccccc	813		

Query Match	10 3%	Score 122:	DB 104:	Length 362:
Best Local Similarity	66.6%	Fred.	No.	1.7e-20:

Matches	241:	Conservative	0:	Mismatches	55:	Indels	66:	Gaps	2
OY	298	caattacacacgctcgcttgtaagttctcaacggagatctgcagactcgcaccaagccctagctcg	357						
Db	1	CAATTACACACCGCTGTTGAGATTCCTAACCGGCAATGTCACATCGATCAACGCTTTGGTTAG	60						
OY	358	tgaataacgctgctcaaatctcgaattctgtctgattcttcaaaaacctgaagtcgaatctgc	417						
Db	61	AGATTAACGCTTGCTCAATTTGTAATTCGTTTGATTTCTTAACCTACGATCTTAATTTGCC	120						
OY	418	tttaagatttgggggaaagctttagttg--gcttgaigtacataaagatagagcaaga	474						
Db	121	TATTGATATTGGTGGGAAAAAGCTTTGACTGTGGTTGGTTGAAGCTTTAAAGCCCGGGGAGA	180						
OY	475	-----	474						
Db	181	TAAAAACCGGTGGAATTCGACATCGAACCAGATGTTCTGTTCCGTTCCAGTTCCACTCGG	240						
OY	475	--aggctgcgcggaattatactataagatctgctcgcgaaglatgataccggggttcac	531						
Db	241	TTCTGCTGCTGTCGGAGACTTATCTCATGATCATGTTGCCCTGCAGGAGATTGATGCGGGGTTAT	300						
OY	532	aaccgcattcaacccggcctcaagaaagccglatagttcaaacacctgatatcagcat	591						
Db	301	AACGGCGATTACACCGGCTTAATGAAGCGGTTTGTGATACAAACGCGGATATTAACTCGCTT	360						
OY	592	ga 593							
Db	361	GA 362							
RESULT	10								
LOCUS	C99848	457 bp	mRNA	EST	08-OCT-1998				
DEFINITION	C99848	YAC clone C1C8B11 region-specific cDNA	Arabidopsis thaliana						
ACCESSION	C99848								
VERSION	C99848.1	GI:3719205							
KEYWORDS	EST.								
SOURCE	thale cress.								
ORGANISM	Arabidopsis thaliana								
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta								
	Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II								
	Brassicales; Brassicaceae; Arabidopsids.								
REFERENCE	1 (bases 1 to 457)								
AUTHORS	Ito, T.								
TITLE	Regional insertional mutagenesis of genes on Arabidopsis thaliana								
	chromosome V using Ac/Ds transposon in combination with a cDNA								
	screening method								
JOURNAL	Unpublished (1998)								
COMMENT	Contact: Takuya Ito								
	Laboratory of Plant Molecular Biology								
	The Institute of Physical and Chemical Research (RIKEN)								
	3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan								
	Email: taito@rtc.riken.go.jp.								
FEATURES	Location/Qualifiers								
source	1..457								
	/organism="Arabidopsis thaliana"								
	/db_xref="taxon:3702"								
BASE COUNT	112 a 125 c 96 g 122 t								
ORIGIN	2 others								
Query Match	8.2%	Score 97.2;	DB 140;	Length 457;					
Best Local Similarity	74.1%;	Pred. No. 3.2e-14;							
Matches 123:	Conservative	0;	Mismatches 43;	Indels	0;	Gaps	0;		
OY	209	ctgcgcagagaaacaccccaagagctgcgcaatccagacgcgcgaagctgtgctactgaacctc	268						
Db	232	CTCGCTCGCTTACAGGTTTCTCACTTCTGCGCATTTGACGCCGACCTTGCTGCTGAACCTC	351						
OY	269	gactctctctctgcgtctcgaaaaacgcgcgcaatcaaacgcgtcttgaagttctcaacgctc	328						

D<sub>b</sub>      352 GATCTCCCTCCTAGGTTAGAGAAATCGACTCAATTACCTTGGCGTAGGTATTAACGGA 41111  
          |||||  
OY      329 gatttcagactcgacacgcctatgttcgataaacgcgttcaaa 374  
          |||||  
D<sub>b</sub>      412 GATTGTCTGTCGATCAAGCCTCTGTACGTGATTAAGGTTGGTGGA 457  
          |||||

RESULT	11
AM574469	
LOCUS	AM574469 386 bp mRNA
DEFINITION	707092H09.x2 707 - Mixed adult tissues from walbot lab (SK) Zea
ACCESSION	AM574469
VERSION	AM574469.1 GI:7242283
KEYWORDS	EST.
SOURCE	Zea mays.
ORGANISM	Zea mays

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 386)	Walbot, V.	Maize ESTs from various cDNA libraries sequenced at Stanford University	Unpublished (1999)	Contact: Walbot V
			Department of Biological Sciences	
			Stanford University	
			855 California Ave, Palo Alto, CA 94304, USA	
			Tel: 650 723 2227	
			Fax: 650 725 8221	
			Email: walbot@stanford.edu	
			Plate: 707092	row: H column: 09.

FEATURES	Location/Qualifiers
source	1. .386

```

/db.xref="taxon:4577"
/clone_lib="707 - Mixed adult tissues from Walbot lab (SK
) "
/tissue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: tassels, kernel, silk, husk, root, leaf;
Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
differentiated maize tissues from an active Mutator
plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
husk, root, leaf). Unidirectionally cloned."
BASE COUNT
100 a 73 c 122 g 91 t

```

Query Match	6.2%	Score 73.4	DB 92	Length 386
Best Local Similarity	72.5%	Pred. No. 3.3e-08		
Matches	95	Conservative	0	Mismatches 36
				Indels 0
				Gaps 0
QY	764	gtgattcggagtacgaatagaggggttcctcgctgtgttgaaacaagcctccgaattagca	823	
Db	1	GTGATCCGGAGTACGAATAGGGGCGCTCGCGTGTGTCTGAATGACCCGCCGCGCTGCT	60	
QY	824	ggattgcgacattgaagcaagctgcttggaagattgttgagcaagaatagcaataaagctgtg	883	
Db	61	GGCCCTTGCTTCGGAGACAGCGCTACTTGGAGATTGTGCGAAGACGATGCATCGCGCGCTC	120	
QY	884	atgtgtgagaga	894	
Db	121	ATGGTCGAGAGA	131	

RESULT 12	CNS0039G/c	LOCUS	DEFINITION
CNS0039G	1101 bp	DNA	GSS
			03-JUN-1999
			Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR08K10 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.  
AF062021

ACCESSION	AL063921	GI:4941778
VERSION	AL063921.1	
KEYWORDS	GSS.	
SOURCE	Fruit fly.	
ORGANISM	Drosophila melanogaster	

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (chance 1 + 0.101)

REFERENCE  
AUTHORS  
1 (bases 1 to 1101)  
Genoscope.

**TITLE** Direct Submission  
**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 Evry cedex - FRANCE (E-mail : [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

COMMENT  
determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila genome Project (BDGP).

melanogaster genome using these BACs. For further information  
the BDGP is constructing a physical map of the *Drosophila*  
genome using these BACs. For further information  
please see  
<http://www.fruitfly.org> and the non-research

please see <http://www.itallil.org> the Boor Itosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Dieter deToni's laboratory in the Department of

tion. The library is named RPCT-98 and was constructed by partial

ECORI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain V2: CN bw<sup>sp</sup>, the same strain used for the BDGP's

PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	Location/Qualifiers
source	1. .1101

```
/organism="Drosophila melanogaster"
/db_xref="taxon:722"
```

```

/clone_11b="RPCI-98"
/clone="BACR08K10"
/clone_11a="BACR08K10"

```

BASE COUNT	/note="end : 1E5"			
ORIGIN	201 a	64 c	131 g	202 t
				503 others

Query Match	6.2%	Score 72.8;	DB 190;	Length 1101;
Best Local Similarity	18.1%	Pred No	5.4e-08.	

Matches 112; Conservative 253; Mismatches 255; Indels 0; Gaps 0;

563 ttgattacaacacctatattactgcattgagatgcatagataggttacagcttgctt 622

Db 1016 DTAWTWRTAWRADWAGRDRGAGKKRDRDATDADGAGRRDGGRRKKDKKDRKDGDDKKG 957

[illegible]

Dd 956 GKKKKA<sup>AKA</sup>KWAT<sup>IKW</sup>DDWD<sup>MD</sup>DKDK<sup>WD</sup>GAK<sup>DR</sup>KAD<sup>DD</sup>DGAG<sup>DK</sup>DDDG<sup>GK</sup>GAD<sup>DD</sup>DT<sup>DG</sup> 897

683 agggggtgaagatagatgtctcaagtccttgatgttcaagagatgttgggatgtcatgttgg 742

00: 743 896 DKDDDDKDKRMDPKAKGTSDATMAWAAT DMMWMMGNDADADMM TMDAADDDMMADDRMDA 837

[illegible][illegible]

776 TTTTTRDTDDWKKWKTPTWTBKAADRTTMDRBDDBDDDBPACTAGCPRKBPRTWKBPBKBBTTPWAD 717

863 caagatagcctgacatgctggaagaaacctaataaaagagcaatcttctcga 922

716 DADADDATDARRRRGGDGDAGAGKKTGRRRRRRDRATWDRPDADAWAAWTTTTDDT 657

OY 923 ttttttgaagtttaqtatcgaatlcqttgaatcqtltcaqtlbqqtltttttqtaqaa 982

```
Db      656 DDMDKRDRRRKGARRRRRTTARAANWWTWKAKWDMAKMDWKTIPADRWDRRAADTTDARK 597
```





Qy 961 ttctctctcttgagcgc 978  
 |||  
 Db 961 ttctctctcttgagcgc 978

## RESULT 2

LOCUS AB009056 65958 bp DNA PLN 09-AUG-2000  
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MZF18.  
 ACCESSION AB009056  
 VERSION AB009056.2 GI:9758219  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_1lb:Mitau1 P1

## ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsids.

## REFERENCE

1 (sites)  
 Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asanizu,E., Miyajima,N.  
 and Tabata,S.  
 and Tabata,S.

## TITLE

Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
 Sequence features of the regions of 1,456,315 bp covered by  
 nineteen physically assigned P1 and TAC clones

DNA Res. 5 (1), 41-54 (1998)

## JOURNAL

98290546  
 2 (bases 1 to 65958)

## MEDLINE

Submitted (27-NOV-1997) to the DDBJ/EMBL/GenBank databases.  
 Yasukazu Nakamura, Kazusa DNA Research Institute, Department of  
 Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:yasukazu@kazusa.or.jp, Tel:81-438-52-3935,  
 Fax:81-438-52-3934)

## AUTHORS

On Aug 9, 2000 this sequence version replaced gi:2656032.

## COMMENT

Address for correspondence: kazusaka@kazusa.or.jp  
 For the latest information on annotation of this clone, please see  
 http://www.kazusa.or.jp/kaos/cgi-bin/agd-graph.cgi?c=MZF18  
 Genes with similarity to proteins in the databases are described in  
 'product' or 'note' qualifiers. Genes that have no significant  
 protein similarity are described as 'unknown protein'.  
 The software programs used to predict genes include: Graal  
 (Informatics Group, Oak Ridge National Laboratory,  
 http://combio.ornl.gov/Graal-1.3/),  
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
 NetGene2 (S.M. Hedsgaard, et al., CBS, Technical University of  
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
 SplicePredictor (Volker Brendel, Stanford University,  
 http://gremli.zool.iastate.edu/cgi-bin/sp.cgi).  
 Genes encoding tRNAs are predicted by tRNAscan-se  
 (Sean Eddy, Washington University School of Medicine, St. Louis,  
 http://genome.wustl.edu/eddy/tRNAscan-se/).  
 This sequence may not be the entire insert of this clone. It may be  
 shorter because we remove overlaps between neighboring submissions.  
 The 5' clone is M001 and the 3' clone is M188.

## FEATURES

## source

1. 65958  
 /organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/chromosome="5"

/clone="MZF18"

/clone\_1lb="Mitau1 P1"

join(8240..8326,8416..8695,8789..9161,9246..9406,  
 9571..9698,9776..10024,10109..10399)

## CDS

/note="gene\_id:MZF18.17"  
 /codon\_start=1  
 /pseudo  
 /evidence=not\_experimental  
 /product="non-LTR retroelement reverse transcriptase-like  
 protein"

## CDS

join(8240..8326,8416..8695,8789..9161,9246..9406,  
 9571..9698,9776..10024,10109..10399)

/note="gene\_id:MZF18.16"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="(+)-delta-cadinene synthase (d-cadinene  
 synthase)"

/protein\_id="BAB08719.1"  
 /db\_xref="GI:9758220"

/db\_xref="GI:9758220"

/translation="MGSEVNRPLADPANIWEEDLPSPKSLDGETFEKESHTLKEA  
 VKRPNSSKANPIENIKFIDALCLGVSEKDVLOLQKSPEDLPDPVNRQECG  
 LITVGIITFOYPROGFKLSADVEKFEKENDCKFGHLITDYGMLSLYEAOKSTHCE  
 DIIIDEALAFSRSHLEISRSRPHLARIRKALNHPYHKGISRIETQVISTYEEES  
 CDPLLEFARKIDFLQIILHREELACVTRWHHEKPEKSVYTRHRTTEALWELGTY  
 FEPQYSGARVITTMALILEFALDQDEIPIPDSMKPIYNTVFEYDQIDELEWEGRS  
 GCGFHLKSLQKTANGYMOEAKMLKTKPIIPFDEYKENALISGVYALIAATFVMTD  
 VAKLDAFEMLSHPKIRIVASEIISRFDDISYEFEKREHVAITGICQVMOFGVSK  
 RAYVMGNIVSDAMKDLNQLMRHVPFPLMLVNLNSRIDVFYRQDVTNPKLL  
 KENHVSLLIETIP"

11597..12883  
 /note="gene\_id:MZF18.15"

/codon\_start=1

/evidence=not\_experimental  
 /product="acetyl-CoA:benzylalcohol acetyltransferase-like  
 protein"

/protein\_id="BAB08720.1"

/db\_xref="GI:9758221"

/translation="MKLELSEVIRKPAHPNHILQTLSTLSEDFQLPSTTVSAIFFYD  
 HSNQEDIORLKSLSQTLSEFLYPLAGQIKDQVTVHCNDGALFTEARKEIFLSDFLRN  
 PSDADLQKFIYSPDHADPEFTWPLHNVIFPKDGFVAVSAGKIDDAISLSTFVC  
 SWTKASGVDVTDPEFGADVPADISIEFPLLVHETSKTRKRFVGLMIETLK  
 NRASSGRVPQATRIESTITALLRCMTFAGSKSGKVEFATOTMDRPPVSSLLP  
 HKAGNRPPLPLKESSEKWEIETETVKLOKTOELNELLRNQSEDDKSSVAKERI  
 ASNALSLCEISPEMETRYANSKCRMSFTYKANKFCGRVWVAPDSVDKTVQVYLDKSD  
 SEGVARTLTETDMARKHDSLVIVTSPSILI"

complement(join(13438..13580,13695..14419,14516..14753,  
 14956..15283,15400..15634,15951..16167,16240..16311,  
 16400..16553))

/note="gene\_id:MZF18.14"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08721.1"

/db\_xref="GI:9758222"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.13"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08722.1"

/db\_xref="GI:9758223"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.12"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08723.1"

/db\_xref="GI:9758224"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.11"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08724.1"

/db\_xref="GI:9758225"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.10"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08725.1"

/db\_xref="GI:9758226"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.9"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08726.1"

/db\_xref="GI:9758227"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.8"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08727.1"

/db\_xref="GI:9758228"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))

/note="gene\_id:MZF18.7"

/codon\_start=1

/evidence=not\_experimental  
 /product="PRO2 homolog"

/protein\_id="BAB08728.1"

/db\_xref="GI:9758229"

/translation="MGNRSVLKTLMLVVLFLQWILVIMINISLNLFKSKATPKSLYN  
 TTYGPOSTNLVLTLPMPFIATVLSVYLHINKLSOGLYECRMKAKIRGRIYIM  
 NTPGIVATLSTLFLVALLANSLVYLYLSYVHLHNDNAKIMQAKIFALRI  
 GYCNICAGFLPEPRTASTITLPLGLTSESSIKYHILGHSNCEFLVHTVFLIY  
 AMINKMETPMPNRYVYNLGTITAMVIGIMWYSLSPFRKKEIFEPYHNLGIY  
 IIVYIIVHGDMPCHILPNIPLFLIDRLRLPLOSIRKSLVSARILPNDLELTFSKT  
 PGLHTPTSLIFLHPSISKIOWHPFTITSSNLKDLTVSVIRQGSNTOKLYTHLS  
 SSIDLSLEVSTGPPVGNPDVSRHNSLILVSGSGITPFIIVIRLISOSONKSTKLP  
 DVLLVGSFKRHDLAFDLIFPLDMSADISRLNRIEAYITREDKKETPDHRLQ  
 TKMFPQPLDPSIPVLGNPLNGLGVILSVFVLILGIVTYIYIPVDHNGSI  
 VNFYSRGIMDFGASCIETSSVYELMRKONKGGDEKQVOSVGEVOTQPSGSI  
 WPHGERLESPVPOSIYOATQSVHSGSNPLKILLLEBGSDDVGVNMGCPKMHNEV  
 AKISSGGLAKMLHFAISFNW"

complement(join(20836..20978,21097..21821,21910..22147,  
 22283..22610,22727..22973,23265..23456,23532..23628,  
 23708..23861))



CDS  
 KLLQTKRFPQPLDPSIPVLPNNFLLGVVITSSFYWELLIGITVIRYIYVPVDHN  
 TGSITNFTYRLEWVFLGVCIFISSIIIFMRKKEKNGEDSGSKVOVSVEQPTPS  
 SPGFMHGERHELEWVPLQSVIATVSHFGSKPNNKILFEFGEDSGVWVCGRPKMS  
 RHEVAKISSGLAKNLEHSAISFNM  
 complement(Join(25651..25974,26145..26386,26718..26876,  
 27106..27216,27399..27625,27695..27963))  
 /note="contains similarity to unknown protein  
 dbj|BAA0637.1  
 gene\_id:MZF18.12"

CDS  
 /evidence-not\_experimental  
 /protein\_id="BAB08723.1"  
 /db\_xref="GI:9758224"  
 /translation="MALLTVSPNPPLSPITSSILVSPSRHLLSSITIRKSPPT  
 LTPHRIYRRRRYVVSOLRPLSPDHSOMALPAAGAGVMSKRTIGSVAG  
 ALSTRLGLAASNLRIPEPTSGVGFEPILPHIPLIPADLRITRSGSLLA  
 FLISVATVGVAVAFVLPKSLGPDMMKTAALMGSTIGSLFVVAISEALQSPS  
 VVAGVAVDNVICALHEWVLPALASKIPETASASSPDMDTKDKLDEKRVSTSI  
 ALVSFLICKAKAITLTTLTKIGVMPVATATVLAISFPDFNSLAPSAETLSIL  
 MOVETTLGATGSSVNVNTAPSIPLFAIQVHVAVLVIGLKLCDMLLLASN  
 ANIGPTTACAMATKAGTSLVPGILSGVEGSIATFGLGCGVFLKRL"  
 29181..31655  
 /note="gene\_id:MZF18.11"  
 /evidence-not\_experimental  
 /product="receptor-protein kinase-like protein"  
 /protein\_id="BAB08724.1"  
 /db\_xref="GI:9758225"  
 /translation="MAFPIQLQTLFFECPLHLHSAFAPTDNYLINSNGNTSF  
 FTRRFLDSSSPGSSFLSTDRSISIDNPSPDSVFLNTRARVPVGSYFQVTK  
 GTHFRLHAPKAPRNLSAKFVILNGFVINSFSSVSVVEFLKIDPVLVEI  
 SFLPKASGFGVNAVEVSAKDYIMOGTKLVIPNSQIINSLSOVLERYHIN  
 GSKRLPNDITMRTVVDNYLLRAARARMTSPYONGATREAPDNYVTA  
 OEMDRNOELARENISNGFOVDERVYLIVLHCDIYSSLSNOLYEVFVNEYVRI  
 KDVDSTLTHLAPLYIDFAVSDRSMPLISVGPSTSPARNALINCYETRI  
 LSPVSEVSGKRVVLYVGSVLDGEVFLSLPFLSYLCLCRKNNKTSSESTQPT  
 LRFREGSSNRTERTVSSGHTLKLISFAQSQGNNDRLVIGVGFQGFQSL  
 KDMTVAAKRGSPGSRQGLPEFLSETTILSKIRHRLVSLVGCESQSMILVYEDM  
 KGPLSKHLYGSTNPPLSKORLEVCIGARGAHLHTGSSOGIIRHDIKSTNLIDNN  
 VYAKVADGLSISGCPIDETHYSTGVSGFYLDPEYRQOGLTSDVSYFEVLE  
 VYCARVADPLVLEOVNLAEMAIEMORGMDOIVDPIMADEIKCSLKEFEVLEK  
 CCADYGVDRPTIGVILMNLHVLLOESGPLNIPEDYDVTDPFRAAGLSNGSIE  
 RYDGCSTGISSTQVFSQLMNAGR"  
 32002..32982  
 /note="dbj|BAA90261.1  
 gene\_id:MZF18.10"  
 /evidence-not\_experimental  
 /product="septum site-determining MIND"  
 /protein\_id="BAB08725.1"  
 /db\_xref="GI:9758226"  
 /translation="MASLRLSTNOSLLPSSISOKTILSSPRFVNNPSRPIRSV  
 LQFNRPPELAGETPRIVITSGRGVGTTRTANVSLARFESVVALDADGLNRL  
 DLLLLEKRVNTVCVINGDCRLDQALVRDRKWSNFEILCTSKPSKPLMGFGKAL  
 EMLVDAKTRRPGSPDFIIDCPAGIDAFITAITPANAVALVITTDITLRLRADVT  
 GLLECDGIRIDIMIVNRRTMDIKEDMNSVLDVOMLSLGLVLPDESEVIRSTNR  
 GPPLVLRKPTIAGLAFEQAAWRLVEODSMKVMVDEEERKRRFFEFQSG"  
 complement(Join(33889..34276,34361..34392,34482..35252,  
 35439..35984,36094..36264))  
 /note="contains similarity to unknown protein  
 gene\_id:MZF18.9  
 PIR|I02898"

CDS

Query Match 100.0%; Score 978; DB 6; Length 65958;  
 Best Local Similarity 100.0%; Pred. No. 6, 9e-277; Indels 0; Gaps 0;  
 Matches 978; Conservative 0; Mismatches 0;

1 atggcgctcctgagatgctcacaacacatcaatcctctcctccatcctcctc 60  
 Db 32002 ATGGCGCTCTGAGATGCTTCAACGACATCAATCTCTCCCTCATCTCTC 32061  
 61 tcacaaagacttaatacttcacaaagattcgtaataaccctagcagaagatcca 120

Db 32062 TCACAAAGACTTAATATCTTCAACAGATTCTCAATAACCTACGACGAGATCCA 32121  
 121 atacagatcgcctctcattcaattcaatcgaacacccggaactcgcggaagaacgcgctatc 180  
 Db 32122 ATACGATCCGCTCTCAATTAATTAACCGAAACCGGAACCTCGCGGAGAAACCCCGCTATC 32181  
 181 gtgcgtatcacctcccggaagaagcggtgtgttgaaagacgaacacccggaagtgctgc 240  
 Db 32182 GTGCGTATCACCTCCGGAAGAGCGGTGTGTGGAAGACGACACACCGCAATGTGCGT 32241  
 241 cctctcctcgcctcgtatcaggtlctcagttgctgcacatcgacgcgaacctgtgtccgt 300  
 Db 32242 CTCTCTCTCCGCTGTACGGTGTCTCTCAGTTGTCCCATTTGACGCCGACCTTGCTCTCGT 32301  
 301 aacctgacatccctcccaagggttgtagaagatcgaatcaatcacctcgtcgcgaagtata 360  
 Db 32302 AACCTGATCTCTCCCTCAAGGTTAGAGAAATGATCAATTAATTAATTAATTAATTA 32361  
 361 aacggaatgtcgtcgtcgcgaacagctcgttacgtaagaagcgtcgcgaatccgaa 420  
 Db 32362 AACGGAATGTCTCTCGATCAAGCTCTGTAGTGTAGCGTGTGTGTGCAATTTGCA 32421  
 421 ttgcatagtatcatcacaactagatcgaacatccgcgaatgagatcgtgtgtaagaacttg 480  
 Db 32422 TTGCTATGTAATCTTAACCTAGATCGAATCTTCGATGCGATTTGCTGTAAAGCATTG 32481  
 481 gaatgctgtgtgagatcgttgaaactagacgcgaaggttaccggaattcatcatc 540  
 Db 32482 GAATGCTGTGTGATGTGCTGTAAGAACTAGACCGGAAGGTTACCGGATTTATCATCTATC 32541  
 541 gatgtcctcgaagaacatcgaatcgcgaatcacaacccgcatctacccgcgaatgaaga 600  
 Db 32542 GATGTCTCTGAGAAATCGAATCGACCGGATTCATTAACCGCATTAATCCGCGCAATGAACA 32601  
 601 gttctgtaacaactccggatataacacgcgttgaaggatcgttaaggatgaagggtttg 660  
 Db 32602 GTTCTGTAAACACTCCGGATATTAACACGCTTAAGCGATGCTGATAGGTTACGGGTTTG 32661  
 661 ttgaatcgaatcgaatcgaatcgaatataagaatgtatgtgaacgaagatgaagaactatg 720  
 Db 32662 TTGAATAAGCGATGGAATCAGAGATATAAGATGATGTAAGACAGATGAGAACTATATG 32721  
 721 attaaaggaagatgatgatgtcagatgtatgaatgtgaatgtgaacgaatgtgaatgt 780  
 Db 32722 ATTAAGAGAGAGATGATGATGTCAAGTGTGAATGTGCGAGACATTTGGATGTGCATG 32781  
 781 ctgtgttaattcctgaagatcctgaagttatcgaacacgaatcgaagggttccgctc 840  
 Db 32782 CTGTGTAAATTCCTGAAGATTCGAGCTTATTCGAACACGAAATCGAGGCTTCCGCTT 32841  
 841 gttctgataagactcccaagcctcgcgaatcgttgaggtttgacgaagcgcgttgagactc 900  
 Db 32842 GTTCTGAATTAAGCTCTCACTACGCTTCGGGATTTGGCGTTTGACACAGCGGCTTGAGACTC 32901  
 901 gttagcaagatagatgaagagctgtatgtgtgaagagaagaacctaaagacgtgcttc 960  
 Db 32902 GTTAGCAAGATATATAGAGAGCTGTATGTGTGAGAGAAACCTTAAGAAAGCTGGCTTC 32961  
 961 ttcctctccttgagcgc 978  
 Db 32962 TTCCTCTTCTTGGCGG 32979

RESULT 3  
 AF251019 1164 bp mRNA 06-SEP-2000  
 LOCUS AF251019  
 DEFINITION Tagetes erecta mind mRNA, complete cds.  
 ACCESSION AF251019  
 VERSION AR251019.1, GI:9971821  
 KEYWORDS  
 ORGANISM African marigold.  
 Tagetes erecta  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

